GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:30:20 ; Search time 7544.92 Seconds
(without alignments)
11245.517 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagttacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2071.6	99.9	2074	6	AX468604	AX468604 Sequence
2		2071.6	99.9	2074	9	AF131884	AF131884 Homo sapi
3	c	1395	67.3	158357	9	AL365434	AL365434 Human DNA
4	c	1395	67.3	160350	2	AC074094	AC074094 Homo sapi
5	c	664.8	32.1	50111	9	AL590622	AL590622 Human DNA
6		367.8	17.7	2358	6	AX468603	AX468603 Sequence
7		367.8	17.7	142902	2	AC119234	AC119234 Mus muscu
8		365	17.6	229640	2	AC105469	AC105469 Rattus no
9		365	17.6	238344	2	AC097115	AC097115 Rattus no
10		340.8	16.4	723	10	AF478692	AF478692 Mus muscu
11		228	11.0	1901	6	AX322775	AX322775 Sequence
12		228	11.0	1901	6	BD094076	BD094076 Shear str
13		228	11.0	1901	9	HSRNACINP	X83703 H.sapiens m
14		228	11.0	1901	11	G28603	G28603 human STS S
15		154.4	7.4	1940	4	AF131883	AF131883 Oryctolag
16		144.4	7.0	1889	6	AX281749	AX281749 Sequence
17	c	133.6	6.4	838	11	BV062893	BV062893 S212P6025
18	c	102.2	4.9	210651	2	AC026715	AC026715 Homo sapi
19	c	101	4.9	160901	2	AC120060	AC120060 Rattus no
20		101	4.9	235419	2	AC095979	AC095979 Rattus no
21	c	100.6	4.9	38476	9	U73024	U73024 Homo sapien
22		99	4.8	116803	2	AC097718	AC097718 Homo sapi
23	c	99	4.8	157665	9	AC104695	AC104695 Homo sapi
24	c	98.8	4.8	87402	9	HSJ437M21	AL049758 Human DNA
25		98.4	4.7	186870	9	AC026471	AC026471 Homo sapi
26		98.4	4.7	227074	2	AC122863	AC122863 Mus muscu
27	c	98.2	4.7	51529	9	AL591291	AL591291 Human DNA
28	c	98.2	4.7	148259	10	AC104099	AC104099 Mus muscu
29	c	98	4.7	172260	9	AC068946	AC068946 Homo sapi
30	c	97.8	4.7	152696	2	AC129985	AC129985 Homo sapi
31	c	97.8	4.7	153053	9	AC090711	AC090711 Homo sapi
32	c	97.8	4.7	163433	2	AC079993	AC079993 Homo sapi
33	c	97.8	4.7	179040	2	AC022732	AC022732 Homo sapi
34		97.6	4.7	51370	2	AC084305	AC084305 Homo sapi
35	c	97.6	4.7	218309	2	AC137376	AC137376 Rattus no
36		97.6	4.7	249682	2	AC096394	AC096394 Rattus no
37	c	97.4	4.7	5404	9	HSM804865	AL833552 Homo sapi
38	c	97.4	4.7	34917	9	AL390211	AL390211 Human DNA
39		97.2	4.7	34488	6	BD093713	BD093713 Human 1p3
40		97.2	4.7	90289	2	AP003164	AP003164 Homo sapi
41		97.2	4.7	121162	6	AX039602	AX039602 Sequence
42		97.2	4.7	171612	9	AL358013	AL358013 Human DNA
43	c	97	4.7	124321	9	HS417G15	AL009174 Human DNA
44		97	4.7	163338	9	AL162426	AL162426 Human DNA
45	c	96.8	4.7	142952	2	AC040909	AC040909 Homo sapi

ALIGNMENTS

RESULT 1
AX468604
LOCUS AX468604 2074 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 2 from Patent WO0246220.
ACCESSION AX468604
VERSION AX468604.1 GI:21901403
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and uses thereof

```
  JOURNAL   Patent: WO 0246220-A 2 13-JUN-2002;
            Aventis Pharma S.A. (FR) ; The Regents of The University of
            California at San Diego (US) ; Benoit, Patrick (FR)
FEATURES             Location/Qualifiers
     source          1. .2074
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      612 a    469 c    416 g    572 t      5 others
ORIGIN

  Query Match             99.9%;  Score 2071.6;  DB 6;  Length 2074;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2074;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60

Qy         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120

Qy        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180

Qy        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240

Qy        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300

Qy        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360

Qy        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420

Qy        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480

Qy        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540

Qy        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600

Qy        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660

Qy        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720

Qy        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780

Qy        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840

Qy        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900

Qy        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960

Qy        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020

Qy       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080

Qy       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140

Qy       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200

Qy       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260

Qy       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320

Qy       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380

Qy       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440

Qy       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500

Qy       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560

Qy       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620

Qy       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680

Qy       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740

Qy       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800

Qy       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860

Qy       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920

Qy       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980
```

```
Qy       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040

Qy       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||||||||||||
Db       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074


RESULT 2
AF131884
LOCUS       AF131884                2074 bp    DNA     linear   PRI 09-FEB-2000
DEFINITION  Homo sapiens cardiovascular-specific cardiac ankyrin repeat protein
            (CVARP) gene, 5'-flanking region and exon 1.
ACCESSION   AF131884
VERSION     AF131884.1  GI:6940841
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2074)
  AUTHORS   Aihara,Y., Kurabayashi,M., Tanaka,T., Sekiguchi,K., Tomaru,K.,
            Kanai,H., Takeda,S. and Nagai,R.
  TITLE     Human CVARP 5'-flanking region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2074)
  AUTHORS   Aihara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1999) Second Department of Internal Medicine,
            Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi
            371-8511, Japan
FEATURES             Location/Qualifiers
     source          1. .2074
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1. .1832
                     /note="5'-flanking region"
     gene            1833. .>2074
                     /gene="CVARP"
     mRNA            1833. .>2074
                     /gene="CVARP"
                     /product="cardiovascular-specific cardiac ankyrin repeat
                     protein"
                     /note="CARP"
     exon            1833. .2074
                     /gene="CVARP"
                     /number=1
BASE COUNT      612 a    469 c    416 g    572 t      5 others
ORIGIN

  Query Match             99.9%;  Score 2071.6;  DB 9;  Length 2074;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2074;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60

Qy         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120

Qy        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180

Qy        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240

Qy        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300

Qy        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360

Qy        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420

Qy        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480

Qy        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540

Qy        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600

Qy        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660

Qy        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720

Qy        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780

Qy        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840

Qy        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900

Qy        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960

Qy        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020

Qy       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080

Qy       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTTBTCATATATATATA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTTBTCATATATATATA 1140

Qy       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200

Qy       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260

Qy       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320

Qy       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380

Qy       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440

Qy       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500

Qy       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560

Qy       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620

Qy       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680

Qy       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740

Qy       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800

Qy       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860

Qy       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920

Qy       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980

Qy       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040

Qy       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
                |||||||||||||||||||||||||||||||
Db       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074


RESULT 3
AL365434/c
LOCUS       AL365434              158357 bp    DNA     linear   PRI 21-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-236B18 on chromosome 10,
            complete sequence.
ACCESSION   AL365434
VERSION     AL365434.13  GI:21540024
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158357)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 23, 2002 this sequence version replaced gi:12191663.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-236B18 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .158357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-236B18"
                     /clone_lib="RPCI-11.1"
BASE COUNT    46958 a  32356 c  32839 g  46204 t
ORIGIN

  Query Match             67.3%;  Score 1395;  DB 9;  Length 158357;
  Best Local Similarity   98.4%;  Pred. No. 7.2e-307;
  Matches 1479;  Conservative    4;  Mismatches   10;  Indels   10;  Gaps    7;

Qy          1 CTGCAGCAAGTTACTTAATGTTT-TTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATT 59
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1502 CTGCAGCAAGTTACTTAATGTTTCTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATT 1443

Qy         60 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAA 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAA 1383

Qy        120 ACATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 ACATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCA 1323

Qy        180 GGCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 GGCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTT 1263

Qy        240 AGATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAA 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 AGATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAA 1203

Qy        300 AATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 AATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 1143

Qy        360 TGAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 TGAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAA 1083
```

```
Qy       420 GAAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTA-- 477
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      1082 GAAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTAATTAAC 1023

Qy       478 CCCCCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGG 537
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1022 CCCCTTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGG 963

Qy       538 TCCCTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGA 597
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       962 TCCCTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGA 903

Qy       598 AACACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTG 657
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       902 AACACCATCCACTGACTGAGCATTCAAGGGGGAAGAGGAGAATGGCAGCCACATTTGTTG 843

Qy       658 ATTGGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACA 717
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       842 ATTGGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACA 783

Qy       718 CTTCCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAA 777
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||:| |:|||
Db       782 CTTCCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTTTCTTTTACTGAAAAAAA 723

Qy       778 CCCAGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAA 837
            |||||||||||||||||||||||| | ||||||||| ||||||||||||||||||||||||
Db       722 CCCAGTTTTTCCTGAAACTATAAAAAGA-CCCCAGTATTTTTACATAATTTACACCTCAA 664

Qy       838 AGATTAGAAACCAGAAATAGAGACC-TTTTCAACCCTTCCGGAAGCAAAGTGCATTATCC 896
            |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       663 AGATTAGAAACCAGAAATAGAGACCTTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCC 604

Qy       897 CTCCAGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAA 956
            |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db       603 CTCCAGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTGAAATTCAA 544

Qy       957 GATGATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGG-CCAGGGAT 1015
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       543 GATGATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCCAGGGAT 484

Qy      1016 CTGTATTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAG 1075
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       483 CTGTATTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAG 424

Qy      1076 CTCAATGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATAT 1135
            |||||||||||||||||||||||||||||||| ||||||||||||:|||||||||||||||
Db       423 CTCAATGACCTAATCAGAGTCCTGCCATTGCTAATAACTGGTCTCATTTTTTTCATATAT 364

Qy      1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
            |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       363 ATATATAGTATTTTTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 304

Qy      1196 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 1255
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       303 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 244

Qy      1256 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATT 1315
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       243 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATT 184

Qy      1316 TTGAAGTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGC 1375
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 TTGAAGTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGC 124

Qy      1376 TA---ATGGCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTA 1432
            ||   |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       123 TAATTATGGCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATTTCTTA 64

Qy      1433 TAATGATGGTAAGGTGGCATGTCATGGGGCCTATTTAG-CCCAGACATCACTCCAAAGAA 1491
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        63 TAATGATGGTAAGGTGGCATGTCATGGGGCCTATTTAGCCCCAGACATCACTCCAAAGAA 4

Qy      1492 TTC 1494
```

```
            |||
Db          3 TTC 1
```

```
RESULT 4
AC074094/c
LOCUS       AC074094              160350 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 10 clone RP11-236B18, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC074094
VERSION     AC074094.3  GI:9958197
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160350)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160350)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:9665205.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0236B18
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 149545 bases at least Q40
            Consensus quality: 152532 bases at least Q30
            Consensus quality: 153924 bases at least Q20
            Insert size: 151000; agarose-fp
            Insert size: 158750; sum-of-contigs
            Quality coverage: 4.85 in Q20 bases; agarose-fp
            Quality coverage: 4.71 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2326: contig of 2326 bp in length
            *     2327     2426: gap of unknown length
            *     2427     7519: contig of 5093 bp in length
            *     7520     7619: gap of unknown length
            *     7620    15286: contig of 7667 bp in length
            *    15287    15386: gap of unknown length
            *    15387    23050: contig of 7664 bp in length
            *    23051    23150: gap of unknown length
            *    23151    33663: contig of 10513 bp in length
            *    33664    33763: gap of unknown length
            *    33764    50133: contig of 16370 bp in length
            *    50134    50233: gap of unknown length
            *    50234    68088: contig of 17855 bp in length
            *    68089    68188: gap of unknown length
            *    68189    88133: contig of 19945 bp in length
            *    88134    88233: gap of unknown length
            *    88234   113397: contig of 25164 bp in length
```

```
            *   113398   113497: gap of unknown length
            *   113498   141732: contig of 28235 bp in length
            *   141733   141832: gap of unknown length
            *   141833   143187: contig of 1355 bp in length
            *   143188   143287: gap of unknown length
            *   143288   144942: contig of 1655 bp in length
            *   144943   145042: gap of unknown length
            *   145043   147286: contig of 2244 bp in length
            *   147287   147386: gap of unknown length
            *   147387   150852: contig of 3466 bp in length
            *   150853   150952: gap of unknown length
            *   150953   153550: contig of 2598 bp in length
            *   153551   153650: gap of unknown length
            *   153651   157257: contig of 3607 bp in length
            *   157258   157357: gap of unknown length
            *   157358   160350: contig of 2993 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160350
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-236B18"
     misc_feature    1. .2326
                     /note="assembly_name:Contig10"
     misc_feature    2427. .7519
                     /note="assembly_name:Contig11"
     misc_feature    7620. .15286
                     /note="assembly_name:Contig12"
     misc_feature    15387. .23050
                     /note="assembly_name:Contig13"
     misc_feature    23151. .33663
                     /note="assembly_name:Contig14"
     misc_feature    33764. .50133
                     /note="assembly_name:Contig15"
     misc_feature    50234. .68088
                     /note="assembly_name:Contig16"
     misc_feature    68189. .88133
                     /note="assembly_name:Contig17"
     misc_feature    88234. .113397
                     /note="assembly_name:Contig18"
     misc_feature    113498. .141732
                     /note="assembly_name:Contig19"
     misc_feature    141833. .143187
                     /note="assembly_name:Contig3"
     misc_feature    143288. .144942
                     /note="assembly_name:Contig4"
     misc_feature    145043. .147286
                     /note="assembly_name:Contig5"
     misc_feature    147387. .150852
                     /note="assembly_name:Contig6"
     misc_feature    150953. .153550
                     /note="assembly_name:Contig7"
     misc_feature    153651. .157257
                     /note="assembly_name:Contig8"
     misc_feature    157358. .160350
                     /note="assembly_name:Contig9"
BASE COUNT    47390 a  32724 c  33135 g  45498 t   1603 others
ORIGIN

  Query Match           67.3%;  Score 1395;  DB 2;  Length 160350;
  Best Local Similarity   92.0%;  Pred. No. 7.2e-307;
  Matches 1482;  Conservative    4;  Mismatches  115;  Indels   10;  Gaps    7;

Qy          1 CTGCAGCAAGTTACTTAATGTTT-TTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATT 59
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     114999 CTGCAGCAAGTTACTTAATGTTTCTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATT 114940

Qy         60 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAA 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114939 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAA 114880

Qy        120 ACATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114879 ACATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCA 114820

Qy        180 GGCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTT 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114819 GGCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTT 114760

Qy        240 AGATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAA 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114759 AGATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAA 114700

Qy        300 AATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114699 AATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 114640

Qy        360 TGAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114639 TGAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAA 114580

Qy        420 GAAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTA-- 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     114579 GAAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTAATTAAC 114520

Qy        478 CCCCCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGG 537
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114519 CCCCTTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGG 114460

Qy        538 TCCCTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGA 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114459 TCCCTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGA 114400

Qy        598 AACACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTG 657
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     114399 AACACCATCCACTGACTGAGCATTCAAGGGGGAAGAGGAGAATGGCAGCCACATTTGTTG 114340

Qy        658 ATTGGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACA 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114339 ATTGGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACA 114280

Qy        718 CTTCCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAA 777
              |||||||||||||||||||||||||||||||||||||||| |||||||||:| |:|||
Db     114279 CTTCCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTTTCTTTTACTGAAAAAAA 114220

Qy        778 CCCAGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAA 837
              |||||||||||||||||||||||| ||| ||||||| ||||||||||||||||||||||
Db     114219 CCCAGTTTTTCCTGAAACTATAAAAAGACC-CCAGTATTTTTACATAATTTACACCTCAA 114161

Qy        838 AGATTAGAAACCAGAAATAGAGACC-TTTTCAACCCTTCCGGAAGCAAAGTGCATTATCC 896
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     114160 AGATTAGAAACCAGAAATAGAGACCTTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCC 114101

Qy        897 CTCCAGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAA 956
              |||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db     114100 CTCCAGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTGAAATTCAA 114041

Qy        957 GATGATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGG-CCAGGGAT 1015
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     114040 GATGATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCCAGGGAT 113981

Qy       1016 CTGTATTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAG 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113980 CTGTATTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAG 113921

Qy       1076 CTCAATGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATAT 1135
              |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db     113920 CTCAATGACCTAATCAGAGTCCTGCCATTGCTAATAACTGGTCTCATTTTTTTCATATAT 113861

Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     113860 ATATATAGTATTTTTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 113801

Qy       1196 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     113800 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 113741

Qy       1256 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATT 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113740 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATT 113681

Qy       1316 TTGAAGTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGC 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113680 TTGAAGTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGC 113621

Qy       1376 TA---ATGGCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTA 1432
              ||   |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     113620 TAATTATGGCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATTTCTTA 113561

Qy       1433 TAATGATGGTAAGGTGGCATGTCATGGGGCCTATTTAG-CCCAGACATCACTCCAAAGAA 1491
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     113560 TAATGATGGTAAGGTGGCATGTCATGGGGCCTATTTAGCCCCAGACATCACTCCAAAGAA 113501

Qy       1492 TTCCAAACAGATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTA 1551
              |||
Db     113500 TTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113441

Qy       1552 CCCAGGGAATAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGAT 1602
                                                              |||
Db     113440 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGTGAGAT 113390
```

RESULT 5
AL590622/c
LOCUS AL590622 50111 bp DNA linear PRI 19-SEP-2001
DEFINITION Human DNA sequence from clone RP11-320F15 on chromosome 10. Contains the gene for ribonuclease P 30kD subunit, the gene for a nuclear protein similar to CARP, ESTs, STSs and GSSs, complete sequence.
ACCESSION AL590622
VERSION AL590622.7 GI:14270159
KEYWORDS HTG; CARP; ribonuclease P.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 50111)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On May 31, 2001 this sequence version replaced gi:14161205.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10
RP11-320F15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-320F15 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-236B18 is at 50012 in this sequence. The true right end of clone RP11-103A2 is at 100 in this sequence.

```
FEATURES             Location/Qualifiers
     source          1. .50111
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-320F15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(1. .98)
                     /note="match: GSS: Em:AQ670367"
     misc_feature    complement(1. .97)
                     /note="match: GSS: Em:AQ544636"
     misc_feature    complement(1. .77)
                     /note="match: STS: Em:G56439
                     match: GSS: Em:AQ311708"
     gene            120. .34469
                     /gene="bA320F15.1"
     mRNA            join(120. .472,3273. .3328,3962. .4018,4428. .4502,7467. .7538,
                     14216. .14305,23179. .23295,23858. .23887,24284. .24321,
                     24717. .24796,28974. .29381)
                     /gene="bA320F15.1"
                     /product="bA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
                     /note="match: cDNAs: Em:U77665 Em:AK004137 Em:U95123
                     match: ESTs: Em:BE572689 Em:BE378859 Em:BE137893
                     Em:AA920575 Em:BG106820 Em:BF681617 Em:BF248001
                     Em:BF031745 Em:AA854455 Em:AI359795"
                     /evidence=not_experimental
     CDS             join(391. .472,3273. .3328,3962. .4018,4428. .4502,7467. .7538,
                     14216. .14305,23179. .23295,23858. .23887,24284. .24321,
                     24717. .24796,28974. .29083)
                     /gene="bA320F15.1"
                     /note="match: proteins: Tr:P78346 Tr:O88796"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
                     /protein_id="CAC70100.1"
                     /db_xref="GI:15717970"
                     /translation="MAVFADLDLRAGSDLKALRGLVETAAHLGYSVVAINHIVDFKEK
                     KQEIEKPVAVSELFTTLPIVQGKSRPIKILTRLTIIVSDPSHCNVLRATSSRARLYDV
                     VAVFPKTEKLFHIACTHLDVDLVCITVTEKLPFYFKRPPINVAIDRGLAFELVYSPAI
                     KDSTMRRYTISSALNLMQICKGKNVIISSAAERPLEIRGPYDVANLGLLFGLSESDAK
                     AAVSTNCRAALLHGETRKTAFGIISTVKKPRPSEGDEDCLPASKKAKCEG"
     repeat_region   689. .800
                     /note="MIR repeat: matches 30. .132 of consensus"
     repeat_region   1443. .1488
                     /note="L2 repeat: matches 2459. .2502 of consensus"
     misc_feature    1761. .2155
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ805524"
     misc_feature    1907. .2434
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ775672"
     misc_feature    1926. .2208
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:B92129"
     repeat_region   2349. .2482
                     /note="L2 repeat: matches 2557. .2696 of consensus"
     repeat_region   2939. .3167
                     /note="MIR repeat: matches 13. .262 of consensus"
     misc_feature    complement(4176. .4649)
                     /note="match: GSS: Em:AQ626500"
     mRNA            join(<4648. .4764,7467. .7538,14216. .14305,23179. .23887,
                     24284. .24321,24717. .24796,28974. .29381)
                     /gene="bA320F15.1"
                     /product="bA320F15.1.2 (putative isoform 2)"
                     /note="match: ESTs: Em:AW939965"
```

```
                /evidence=not_experimental
repeat_region   4733. .4858
                /note="L2 repeat: matches 2601. .2732 of consensus"
repeat_region   4928. .5102
                /note="L2 repeat: matches 2299. .2478 of consensus"
repeat_region   5257. .5439
                /note="L2 repeat: matches 1808. .1971 of consensus"
repeat_region   5440. .5721
                /note="AluSx repeat: matches 7. .293 of consensus"
repeat_region   5722. .6237
                /note="L2 repeat: matches 1200. .1808 of consensus"
repeat_region   6884. .7196
                /note="AluJb repeat: matches 1. .312 of consensus"
repeat_region   8243. .8404
                /note="2 copies 81 mer 81% conserved"
repeat_region   8445. .8541
                /note="HY1 repeat: matches 16. .112 of consensus"
repeat_region   8542. .8583
                /note="HY1 repeat: matches 1. .42 of consensus"
repeat_region   8850. .9130
                /note="AluSx repeat: matches 1. .289 of consensus"
repeat_region   9601. .9741
                /note="MIR repeat: matches 76. .212 of consensus"
repeat_region   10798. .11076
                /note="AluY repeat: matches 1. .291 of consensus"
repeat_region   11086. .11219
                /note="FLAM_C repeat: matches 1. .129 of consensus"
repeat_region   11319. .11491
                /note="MER5A repeat: matches 3. .189 of consensus"
repeat_region   11928. .11984
                /note="L2 repeat: matches 1743. .1801 of consensus"
repeat_region   13755. .13817
                /note="L2 repeat: matches 2684. .2749 of consensus"
repeat_region   15390. .15671
                /note="AluJo repeat: matches 1. .282 of consensus"
repeat_region   15791. .16088
                /note="AluSg repeat: matches 1. .296 of consensus"
repeat_region   16979. .17036
                /note="L2 repeat: matches 2423. .2481 of consensus"
misc_feature    17236. .17755
                /gene="bA320F15.1"
                /note="match: GSS: Em:AQ551877"
repeat_region   17709. .17972
                /note="AluSx repeat: matches 45. .304 of consensus"
repeat_region   19198. .19363
                /note="L1ME repeat: matches 5272. .5435 of consensus"
repeat_region   19489. .19928
                /note="MLT1C repeat: matches 5. .478 of consensus"
repeat_region   21838. .22258
                /note="L2 repeat: matches 2109. .2669 of consensus"
misc_feature    22440. .22946
                /gene="bA320F15.1"
                /note="match: GSS: Em:AQ876596"
mRNA            join(23179. .23295,23858. .23887,24284. .24321,24717. .24796,
                28974. .29072,30647. .30748,31596. .31680,33706. .34469)
                /gene="bA320F15.1"
                /product="bA320F15.1.3 (putative isoform 3)"
                /note="match: ESTs: Em:BE567341 Em:AI292002 Em:BF684192
                Em:BF435407 Em:AI740881 Em:AA651912 Em:BF438915"
                /evidence=not_experimental
misc_feature    24121. .24518
                /gene="bA320F15.1"
                /note="match: GSS: Em:AQ026750"
misc_feature    complement(join(24284. .24321,24724. .24796,28974. .29240))
                /note="match: STS: Em:G24365"
misc_feature    24549. .25020
                /gene="bA320F15.1"
                /note="match: GSS: Em:AQ223714"
misc_feature    24561. .24956
                /gene="bA320F15.1"
                /note="match: GSS: Em:B53566"
repeat_region   25321. .25458
                /note="FLAM_C repeat: matches 1. .133 of consensus"
misc_feature    complement(25417. .25863)
                /note="match: GSS: Em:AQ712354"
repeat_region   26204. .26603
                /note="L1MA2 repeat: matches 5888. .6302 of consensus"
repeat_region   28442. .28516
                /note="L2 repeat: matches 2636. .2710 of consensus"
polyA_signal    29149. .29154
                /gene="bA320F15.1"
polyA_site      29166
                /gene="bA320F15.1"

  Query Match           32.1%;  Score 664.8;  DB 9;  Length 50111;
  Best Local Similarity  99.6%;  Pred. No. 9.2e-141;
  Matches  677;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1396 AAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTC 1455
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db      50111 AAAGAAAAGTGCATTACTGAATGCTTTCAATTTCTTATAATGATGGTAAGGTGGCATGTC 50052

Qy       1456 ATGGGGCCTATTTAG-CCCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTG 1514
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      50051 ATGGGGCCTATTTAGCCCCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTG 49992

Qy       1515 CCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGG 1574
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49991 CCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGG 49932

Qy       1575 ACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGG 1634
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49931 ACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGG 49872

Qy       1635 TGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAA 1694
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49871 TGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAA 49812

Qy       1695 AAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGACTCTTGACAAAT 1754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49811 AAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGACTCTTGACAAAT 49752

Qy       1755 AGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTT 1814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49751 AGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTT 49692

Qy       1815 CTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCA 1874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49691 CTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCA 49632

Qy       1875 GACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGA 1934
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      49631 GACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGA 49572

Qy       1935 GTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGA 1994
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49571 GTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGA 49512

Qy       1995 CCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACAT 2054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49511 CCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACAT 49452

Qy       2055 ACAAGACTCCTTCAGCCAAC 2074
             ||||||||||||||||||||
Db      49451 ACAAGACTCCTTCAGCCAAC 49432


RESULT 6
AX468603
LOCUS       AX468603               2358 bp    DNA     linear   PAT 16-JUL-2002
DEFINITION  Sequence 1 from Patent WO0246220.
ACCESSION   AX468603
VERSION     AX468603.1  GI:21901402
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and uses thereof
JOURNAL Patent: WO 0246220-A 1 13-JUN-2002;
Aventis Pharma S.A. (FR) ; The Regents of The University of California at San Diego (US) ; Benoit, Patrick (FR)

FEATURES Location/Qualifiers
source 1. .2358
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

BASE COUNT 636 a 567 c 542 g 613 t
ORIGIN

Query Match 17.7%; Score 367.8; DB 6; Length 2358;
Best Local Similarity 76.4%; Pred. No. 3.6e-73;
Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;

```
Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
             ||| | | | |||| | || ||||| || | || ||| |||| | | | |
Db       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405
             |||||||||||| ||| ||| || ||| |||||||||||| |||| | |||||
Db       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657

Qy       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465
             |||||||||||||||||||| ||| ||||| ||||| | |||||||||| |||||
Db       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717

Qy       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524
             |||||| |||||||||||||| |||||||||||||||| ||||||||| | | ||
Db       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777

Qy       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTT 1581
             ||| ||| | ||| | || ||| | |||||||||||| | ||| |
Db       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837

Qy       1582 TCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
              | | || || |||| ||||||||  |||||||||||||||  |||  |||||||
Db       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897

Qy       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698
                || ||  | |||||||||||||| | |||||||||||| ||||||||||||  ||||
Db       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955

Qy       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757
              || |||| |||||||||||||| | ||||| |||||||||||||| | ||  || | | ||
Db       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015

Qy       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817
              |||||| |||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075

Qy       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877
              |||||||| |||| | || |      |   |  ||  |||||||| ||||||||||
Db       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130

Qy       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937
              ||| ||||||||||||||||||||||||||||||| |||||  ||  ||||||| || |||
Db       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190

Qy       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997
              | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249

Qy       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056
              |||||||||||  |||||||| ||||||| ||| | | || || ||||||||||
Db       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309

Qy       2057 A 2057
              |
Db       2310 A 2310
```

RESULT 7
AC119234
LOCUS AC119234 142902 bp DNA linear HTG 11-JUN-2003
DEFINITION Mus musculus clone RP24-211P24, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC119234
VERSION AC119234.9 GI:31581760
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 142902)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-211P24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142902)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 142902)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jun 11, 2003 this sequence version replaced gi:31455706.
All repeats were identified using RepeatMasker:

```
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25236
                Center clone name: 211_P_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    61126: contig of 61126 bp in length
            *   61127    61226: gap of 100 bp
            *   61227   142902: contig of 81676 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142902
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-211P24"
                     /clone_lib="RPCI-24 Male Mouse BAC"
BASE COUNT    41376 a  30899 c  29987 g  40473 t    167 others
ORIGIN

  Query Match             17.7%;  Score 367.8;  DB 2;  Length 142902;
  Best Local Similarity   76.4%;  Pred. No. 3.2e-73;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
              ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db      90844 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 90901

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405
              |||||||||||| ||| |||  ||  |||   |||||||||||||| |||| |   |||||
Db      90902 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 90961

Qy       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465
              |||||||||||||||||||| |||  ||||| ||||| | ||||||||||  |||||
Db      90962 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 91021

Qy       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524
              ||||||  |||||||||||||| |||||||||||||||| |||||||| |      | |
Db      91022 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 91081

Qy       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTT 1581
              |||  ||| | |||     |    ||  |||  | ||||||||||||     |  |||  |
Db      91082 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 91141

Qy       1582 TCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
                |  | || || |||| |||||||||   ||||||||||||||  |||   |||||||
Db      91142 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 91201

Qy       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698
                || ||   | ||||||||||||| | ||||||||||| ||||||||||||   ||||
Db      91202 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 91259

Qy       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757
              || |||| |||||||||||||| | ||||| |||||||||||||| | ||  || |  | ||
Db      91260 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 91319

Qy       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817
              |||||| |||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db      91320 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 91379

Qy       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877
              |||||||| ||||  | ||  |       |    |  ||  |||||||||  ||||||||||
Db      91380 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 91434

Qy       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937
              ||| ||||||||||||||||||||||||||||||| |||||  ||  |||||||  || |||
Db      91435 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 91494

Qy       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997
              | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db      91495 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 91553

Qy       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056
              |||||||||||| |  |||||||  |||||||  ||| | | || || ||||||||||
Db      91554 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 91613

Qy       2057 A 2057
              |
Db      91614 A 91614


RESULT 8
AC105469
LOCUS       AC105469               229640 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-140I18, WORKING DRAFT SEQUENCE, 9
            unordered pieces.
ACCESSION   AC105469
VERSION     AC105469.4  GI:25007338
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229640)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
```

```
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229640)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 229640)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23101653.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNAG
                Center clone name: CH230-140I18
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 215126 bases at least Q40
                Consensus quality: 217436 bases at least Q30
                Consensus quality: 218866 bases at least Q20
                Estimated insert size: 214300; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   215467: contig of 215467 bp in length
            *  215468   215567: gap of unknown length
            *  215568   218782: contig of 3215 bp in length
            *  218783   218882: gap of unknown length
            *  218883   220359: contig of 1477 bp in length
            *  220360   220459: gap of unknown length
            *  220460   221614: contig of 1155 bp in length
            *  221615   221714: gap of unknown length
            *  221715   223060: contig of 1346 bp in length
            *  223061   223160: gap of unknown length
            *  223161   224771: contig of 1611 bp in length
            *  224772   224871: gap of unknown length
            *  224872   225886: contig of 1015 bp in length
            *  225887   225986: gap of unknown length
            *  225987   228180: contig of 2194 bp in length
            *  228181   228280: gap of unknown length
            *  228281   229640: contig of 1360 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229640
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-140I18"
     misc_feature    1. .1268
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    4765. .7050
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(7251. .8156)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RWAXF57TJB"
     misc_feature    complement(7252. .8118)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RWAXF57TJC"
     misc_feature    213542. .215467
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    215568. .216791
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    61330 a  47195 c  48097 g  63316 t   9702 others
ORIGIN

  Query Match            17.6%;  Score 365;  DB 2;  Length 229640;
  Best Local Similarity  77.3%;  Pred. No. 1.4e-72;
  Matches  601;  Conservative    0;  Mismatches  150;  Indels   26;  Gaps   12;

Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
              | | | ||| |||||||||||| ||||||||| || || ||  ||| ||||   | |||||
Db     141439 TGTGTCTCTGGACCACTCTGACCCATTTTGAGGTCAATAT--TCCGATTAGCCTTCTGTT 141496

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTAATGGCCAGTGACATCATAAAAGAAAAGTGCA 1408
              ||||| |||    |     |    | | |||||||||||||| ||||||  |||||||||
Db     141497 TTAGAGCAC----ATGCCATGCACTAATTATGGCCAGTGACACCATAAAGTAAAAGTGCA 141552

Qy       1409 TTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTATTT 1468
              |||||||||||||||| ||| |||||||||||||||||| ||||||||||  |||
Db     141553 TTACTGAATGCTTTCAATTTCTCATAATGATGGTAAGGTGGCGTGTCATGGGGCCATTTT 141612

Qy       1469 AGCCC-AGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGCCCA 1527
              |||||  |||||||||||| |||||||  ||||||||||||||||||| |   | | ||||
Db     141613 AGCCCTGGACATCACTCCAGAGAATTCTGAACAGATATAGACAAGTGTCACCAAGACCCA 141672

Qy       1528 GATCCCTTCCCCTCAGGCTGTTTA-CCCAGGGAATAGGATGTCCTGGGACAAGTTTCCC- 1585
                ||||||| | |||||||||  ||  |||||||||||||    |||   | ||
Db     141673 -----CTTCCCCCCGGGGCTGTTTATTCCCAGGAATAGGATGTCCCAAAGCAACACTTCCA 141727

Qy       1586 --CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATATG 1643
                | || || |||| ||||||| |   ||||||||||||||| | | | |||||||
Db     141728 GGCCAACTGGAGTGCTGATAAGCCCAGTTATCAGAAAGATATTGCCGTGAGTGTGATGCA 141787

Qy       1644 TAGGGCATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAA--AAAAAGG 1701
              |  || |  || ||| ||     | |||||||| |||||||||| ||||  | |||||
Db     141788 CAATGCTTGCAC-TTTCCTGATAGGTTAGTCATACGAAAGCTGACAGAGAAGGAGAAAGG 141846

Qy       1702 GCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGACTCTTGACAAATAGGATGA 1761
              |||| |||||||||||| | ||||| ||||||||||  |  ||  || || ||||||||
```

```
Db     141847 GCAGCGATGTGGTGCAATATGAACAGGCAGCTGTCCCTGGCTTCCCGATAAGTAGGATGA 141906

Qy       1762 CTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTTCCTG 1821
              || ||||||||| ||| |||| ||||  |||| ||||||||||||||||||||||||||
Db     141907 CTCGCATTGCTGGGCGGTGTGGTCACTGCCAATGGAATGGCCCTCTCACATTTCTTCCTG 141966

Qy       1822 ATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTG 1881
              |||| || |  || || ||  |||||||| |   |||||||||  |||||||||||||||
Db     141967 ATTCGCACA--CACCACGGCCAGCTTGTCAT---CTCCCTCTTGGGCTTCCCAGACACTA 142021

Qy       1882 AGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTAC 1941
              |||||||||||||||||||||||||||||| |||||  | |  || ||| || |||| ||
Db     142022 AGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACCAGCGGAGGCAGGGGTGTGAC 142081

Qy       1942 TTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGT 2001
              || | |||||||| ||||||||||||||||||| | || ||||||||||  | | || ||
Db     142082 TTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGTCCTAGCTATATAAG-AGGCTGGGGT 142140

Qy       2002 GGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATACA 2057
              |||||||| |  |||||||| |||||||| |||    | || ||  |||||||||| |
Db     142141 GGAGGGGCTCCACAGGGCCAGCTCCAGGGGTTCAGCCACAAGAGGGAAAAACATAGA 142197
```

RESULT 9
AC097115
LOCUS AC097115 238344 bp DNA linear HTG 14-NOV-2002
DEFINITION Rattus norvegicus clone CH230-26A2, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC097115
VERSION AC097115.6 GI:24956605
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 238344)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 238344)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 238344)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 14, 2002 this sequence version replaced gi:22855482.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GGQJ
Center clone name: CH230-26A2
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222569 bases at least Q40
Consensus quality: 226638 bases at least Q30
Consensus quality: 229427 bases at least Q20
Estimated insert size: 233018; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 236546: contig of 236546 bp in length
* 236547 236646: gap of unknown length
* 236647 238344: contig of 1698 bp in length.
FEATURES Location/Qualifiers
source 1..238344
/organism="Rattus norvegicus"
/mol_type="genomic DNA"

```
                     /db_xref="taxon:10116"
                     /clone="CH230-26A2"
     misc_feature    1. .1081
                     /note="wgs_contig"
BASE COUNT    63454 a  47381 c  49225 g  71429 t   6855 others
ORIGIN

  Query Match           17.6%;  Score 365;  DB 2;  Length 238344;
  Best Local Similarity   77.3%;  Pred. No. 1.4e-72;
  Matches  601;  Conservative    0;  Mismatches  150;  Indels   26;  Gaps   12;

Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
             |  | | ||| |||||||||||| ||||||||| || || ||  ||| ||||   | |||||
Db      27068 TGTGTCTCTGGACCACTCTGACCCATTTTGAGGTCAATAT--TCCGATTAGCCTTCTGTT 27125

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTAATGGCCAGTGACATCATAAAAGAAAAGTGCA 1408
             |||||  |||    |     |     |  | |||||||||||||| ||||||  |||||||||
Db      27126 TTAGAGCAC----ATGCCATGCACTAATTATGGCCAGTGACACCATAAAGTAAAAGTGCA 27181

Qy       1409 TTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTATTT 1468
             |||||||||||||||||| ||| ||||||||||||||||||| ||||||||||||  |||
Db      27182 TTACTGAATGCTTTCAATTTCTCATAATGATGGTAAGGTGGCGTGTCATGGGGCCATTTT 27241

Qy       1469 AGCCC-AGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGCCCA 1527
             |||||  |||||||||||| |||||||  |||||||||||||||||| |   | | ||||
Db      27242 AGCCCTGGACATCACTCCAGAGAATTCTGAACAGATATAGACAAGTGTCACCAAGACCCA 27301

Qy       1528 GATCCCTTCCCCTCAGGCTGTTTA-CCCAGGGAATAGGATGTCCTGGGACAAGTTTCCC- 1585
              ||||||| | |||||||||  ||  ||||||||||||||     |||   | ||
Db      27302 -----CTTCCCCCCGGGCTGTTTATTCCCAGGAATAGGATGTCCCAAAGCAACACTTCCA 27356

Qy       1586 --CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATATG 1643
                | || || |||| ||||||| |   |||||||||||||||| | | | |||||||
Db      27357 GGCCAACTGGAGTGCTGATAAGCCCAGTTATCAGAAAGATATTGCCGTGAGTGTGATGCA 27416

Qy       1644 TAGGGCATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAA--AAAAAGG 1701
                |  || |  || ||| ||     | |||||||| |||||||||| ||||  | |||||
Db      27417 CAATGCTTGCAC-TTTCCTGATAGGTTAGTCATACGAAAGCTGACAGAGAAGGAGAAAGG 27475

Qy       1702 GCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGACTCTTGACAAATAGGATGA 1761
             |||| |||||||||||| | ||||| ||||||||||  |   ||  || || |||||||||
Db      27476 GCAGCGATGTGGTGCAATATGAACAGGCAGCTGTCCCTGGCTTCCCGATAAGTAGGATGA 27535

Qy       1762 CTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTTCCTG 1821
             || ||||||||| ||| |||| ||||  |||| ||||||||||||||||||||||||||||
Db      27536 CTCGCATTGCTGGGCGGTGTGGTCACTGCCAATGGAATGGCCCTCTCACATTTCTTCCTG 27595

Qy       1822 ATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTG 1881
             |||| || |  || || ||  |||||||| |   |||||||||  ||||||||||||||
Db      27596 ATTCGCACA--CACCACGGCCAGCTTGTCAT---CTCCCTCTTGGGCTTCCCAGACACTA 27650

Qy       1882 AGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTAC 1941
             |||||||||||||||||||||||||||||||| |||||  | |  || ||| || |||| ||
Db      27651 AGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACCAGCGGAGGCAGGGGTGTGAC 27710

Qy       1942 TTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGT 2001
             || | ||||||||| |||||||||||||||||| | || ||||||||||  | | || ||
Db      27711 TTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGTCCTAGCTATATAAG-AGGCTGGGGT 27769

Qy       2002 GGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATACA 2057
             |||||||| |  |||||||| |||||||| |||   | || ||  |||||||||| |
Db      27770 GGAGGGGCTCCACAGGGCCAGCTCCAGGGGTTCAGCCACAAGAGGGAAAAACATAGA 27826


RESULT 10
AF478692
LOCUS       AF478692                723 bp    DNA     linear   ROD 10-OCT-2002
DEFINITION  Mus musculus cardiac ankyrin repeat protein (Carp) gene, promoter
            region and partial cds.
ACCESSION   AF478692
VERSION     AF478692.1  GI:19110906
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Maeda,T., Sepulveda,J., Chen,H.H. and Stewart,A.F.R.
  TITLE     alpha1-Adrenergic activation of the cardiac ankyrin repeat protein
            gene in cardiac myocytes
  JOURNAL   Gene 297 (1-2), 1-9 (2002)
REFERENCE   2  (bases 1 to 723)
  AUTHORS   Maeda,T., Sepulveda,J. and Stewart,A.F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Cardiovascular Institute, University of
            Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15213, USA
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="FVB"
                     /db_xref="taxon:10090"
     gene            1. .>723
                     /gene="Carp"
     misc_feature    1. .720
                     /gene="Carp"
                     /note="contains promoter and 5' UTR"
     mRNA            <721. .>723
                     /gene="Carp"
                     /product="cardiac ankyrin repeat protein"
     CDS             721. .>723
                     /gene="Carp"
                     /codon_start=1
                     /product="cardiac ankyrin repeat protein"
                     /protein_id="AAL85342.1"
                     /db_xref="GI:19110907"
                     /translation="M"
BASE COUNT      187 a    185 c    180 g    171 t
ORIGIN

  Query Match           16.4%;  Score 340.8;  DB 10;  Length 723;
  Best Local Similarity   74.0%;  Pred. No. 5.2e-67;
  Matches  527;  Conservative    0;  Mismatches  172;  Indels   13;  Gaps    7;

Qy       1352 GAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGTGCA 1408
             |||||||||  ||| | |  ||  |||   |||||||||||||| |||| |   |||||  |
Db             1 GAACACGGTGAGCCTCTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTCGA 60

Qy       1409 TTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTATTT 1468
             |||||||||||||||||| |||  ||||| ||||| | ||||||||||  |||||  |||
Db       61 TTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTT 120

Qy       1469 AGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGCCCA 1527
             |||  |||||||||||||| |||||||||||||||||| |||||||||  |      | ||||
Db       121 AGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCA 180

Qy       1528 GATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGGACAAGTTTCCCCT 1587
                ||| | |||    |   ||  |||  | |||||     ||   |  |  |  |  |
Db       181 TCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGATGTCCCAAAGCAACTTCCCAGCC 240

Qy       1588 AAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATATGTAGG 1647
             || || |    |||||||||   ||||||||||||||| |||   |||||||    ||
Db       241 AACTGGATGCGTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATCGACAGT 300

Qy       1648 GCATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAAAAAAAGGGCAGTG 1707
             || || |  |||  |     | |||||||||| ||||||||||||  ||| |
Db       301 GCTTCGA-TTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAGAGCAGC 359

Qy       1708 ATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGGATGACTTGC 1766
                 ||||||| | ||||| |||||||||||| | ||  || | | ||||||| ||
Db       360 GATGTGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGC 419

Qy       1767 ATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTTCCTGATTCA 1826
             |||||||||| |||| ||||  |||||||||| ||||||||||||||||||||||||||
Db       420 ATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCG 479
```

```
Qy       1827 CATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
             ||||  | ||  |       |   |  ||  |||||||||  ||||||||||||||| |||||
Db        480 CATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCT 534

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
             ||||||||||||||||||||||||| |||||   ||  |||||||  || |||| |||| |
Db        535 GGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGC 594

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
             ||||||||  |||||||||||||||||||  |||| ||||||||||  | | | ||||||||||
Db        595 TTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGG 653

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATACA 2057
             |||  |  ||||||||  |||||||  |||  | | || || ||||||||||  |
Db        654 GGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAGA 705


RESULT 11
AX322775
LOCUS       AX322775               1901 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 19 from Patent WO0192567.
ACCESSION   AX322775
VERSION     AX322775.1  GI:18093755
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Bunk,D., Reuner,B., Beck,J. and Henkel,T.
  TITLE     Novel target genes for diseases of the heart
  JOURNAL   Patent: WO 0192567-A 19 06-DEC-2001;
            Medigene AG (DE)
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match             11.0%;  Score 228;  DB 6;  Length 1901;
  Best Local Similarity   97.6%;  Pred. No. 2.3e-41;
  Matches  242;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
              | |  |||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db          2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
              ||||||||
Db        242 CAGCCAAC 249


RESULT 12
BD094076
LOCUS       BD094076               1901 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Shear stress-responsive DNAs.
ACCESSION   BD094076
VERSION     BD094076.1  GI:22639664
KEYWORDS    WO 0125427-A/37.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1901)
  AUTHORS   Nojima,H., Yoshisue,H., Obayashi,M., Ota,T., Kawabata,A.,
            Sakurada,K., Kuga,T., Sekine,S., Nakamura,Y. and Sugan,S.
  TITLE     Shear stress-responsive DNAs
  JOURNAL   Patent: WO 0125427-A 37 12-APR-2001;
            KYOWA HAKKO KOGYO CO LTD,HIROSHI NOJIMA,HAJIME YOSHISUE, MASAYA
            OBAYASHI, TOSHIO OTA,AYAKO KAWABATA,KAZUHIRO SAKURADA,TETSURO KUGA,
            SUSUMU SEKINE, YUSUKE NAKAMURA,SUMIO SUGANO
COMMENT     OS   Homo sapiens (human)
            PN   WO 0125427-A/37
            PD   12-APR-2001
            PF   02-OCT-2000 WO 2000JP006840
            PR   01-OCT-1999 JP 99P   280976
            PI   HIROSHI NOJIMA,HAJIME YOSHISUE,MASAYA OBAYASHI,TOSHIO OTA, PI
             AYAKO KAWABATA,
            PI   KAZUHIRO SAKURADA,TETSURO KUGA,SUSUMU SEKINE,YUSUKE NAKAMURA,
            PI   SUMIO SUGANO
            PC   C12N15/12,C07K14/435,C07K16/18,C12P21/02,C12Q1/68,A61K38/00,
            PC   A61K39/395,
            PC   A61K48/00,A61P9/10,G01N33/50,G01N33/53
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (250). .(1206).
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match             11.0%;  Score 228;  DB 6;  Length 1901;
  Best Local Similarity   97.6%;  Pred. No. 2.3e-41;
  Matches  242;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
              | |  |||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db          2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
              ||||||||
Db        242 CAGCCAAC 249


RESULT 13
HSRNACINP
LOCUS       HSRNACINP              1901 bp    mRNA    linear   PRI 05-SEP-1995
DEFINITION  H.sapiens mRNA for cytokine inducible nuclear protein.
ACCESSION   X83703
VERSION     X83703.1  GI:793840
KEYWORDS    ankyrin-like repeat; nuclear localisation signal; nuclear protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chu,W., Burns,D.K., Swerlick,R.A. and Presky,D.H.
  TITLE     Identification and characterization of a novel cytokine-inducible
```

```
          nuclear protein from human endothelial cells
  JOURNAL   J. Biol. Chem. 270 (17), 10236-10245 (1995)
  MEDLINE   95247734
   PUBMED   7730328
REFERENCE   2  (bases 1 to 1901)
  AUTHORS   Chu,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1995) W. Chu, Hoffmann-La Roche, 340 Kingsland
            Street, Dept. of Inflammation/Autoimmune Disease, Hoffmann-La
            Roche, Nutley, NJ 07110, USA
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="C-193"
                     /cell_type="endothelial"
                     /tissue_type="skin"
                     /clone_lib="HDMEC cDNA"
     mRNA            1. .1901
     misc_feature    94. .98
                     /note="nuclear localization signal"
     repeat_unit     152. .283
                     /note="ankyrin-like repeats"
     CDS             250. .1209
                     /note="cytokine-inducible expression"
                     /codon_start=1
                     /product="nuclear protein"
                     /protein_id="CAA58676.1"
                     /db_xref="GI:793841"
                     /db_xref="SPTREMBL:Q15327"
                     /translation="MMVLKVEELVTGKKNGNGEAGEFLPEDFRDGEYEAAVTLEKQED
                     LKTLLAHPVTLGEQQWKSEKQREAELPKKKLEQRSKLENLEDLEIIIQLKKRKKYRKT
                     KVPVVKEPEPEIITEPVDVPTFLKAALENKLPVVEKFLSDKNNPDVCDEYKRTALHRA
                     CLEGHLAIVEKLMEAGAQIEFRDMLESTAIHWASRGGNLDVLKLLLNKGAKISARDKL
                     LSTALHVAVRTGHYECAEHLIACEADLNAKDREGDTPLHDAVRLNRYKMIRLLIMYGA
                     DLNIKNCAGKTPMDLVLHWQNGTKAIFDSLRENSYKTSRIATF"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match           11.0%;  Score 228;  DB 9;  Length 1901;
  Best Local Similarity   97.6%;  Pred. No. 2.3e-41;
  Matches  242;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
                | |  |||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db             2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db            62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
                ||||||||
Db           242 CAGCCAAC 249


RESULT 14
G28603
LOCUS       G28603                  1901 bp    DNA     linear   STS 11-JUL-1996
DEFINITION  human STS SHGC-35401, sequence tagged site.
ACCESSION   G28603
VERSION     G28603.1  GI:1408418
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1901)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: GGCATTTTGAAGGCATGG
            Primer B: CCAGATGGATGATCATGAAGG
            STS size: 222
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from X83703
            -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="10"
     STS             1246. .1467
     primer_bind     1246. .1263
     primer_bind     complement(1447. .1467)
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match           11.0%;  Score 228;  DB 11;  Length 1901;
  Best Local Similarity   97.6%;  Pred. No. 2.3e-41;
  Matches  242;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
                | |  |||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db             2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db            62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241
```

```
Qy       2067 CAGCCAAC 2074
              ||||||||
Db        242 CAGCCAAC 249


RESULT 15
AF131883
LOCUS       AF131883               1940 bp    mRNA    linear   MAM 30-NOV-1999
DEFINITION  Oryctolagus cuniculus CARP mRNA, complete cds.
ACCESSION   AF131883
VERSION     AF131883.1  GI:6478316
KEYWORDS    .
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 1940)
  AUTHORS   Aihara,Y., Kurabayashi,M., Arai,M., Kedes,L. and Nagai,R.
  TITLE     Molecular cloning of rabbit CARP cDNA and its regulated expression
            in adriamycin-cardiomyopathy
  JOURNAL   Biochim. Biophys. Acta 1447 (2-3), 318-324 (1999)
  MEDLINE   20011295
   PUBMED   10542334
REFERENCE   2  (bases 1 to 1940)
  AUTHORS   Aihara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1999) Second Department of Internal Medicine,
            Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi
            371-8511, Japan
FEATURES             Location/Qualifiers
     source          1. .1940
                     /organism="Oryctolagus cuniculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9986"
     CDS             219. .1178
                     /codon_start=1
                     /product="CARP"
                     /protein_id="AAF13817.1"
                     /db_xref="GI:6478317"
                     /translation="MMVLKVEELVTGKKSSNGETGEFLPDDFRDGQYEAAVTSEKQED
                     LKTLPAHHVSLAEQQWEREKQLEAELKKKKLEQRSKLENLEDLEIIIQLKKRKKYRKT
                     KVPVAKEPEPEIITEPVDVPRFLKAALENKLAVVEKFLSDQNNPDVCDEYKRTALHRA
                     CLEGHLAIVEKLMEAGAQIEFRDMLESTAIHWACRGGNLEVLKLLLNKGAKISARDKL
                     LSTALHVAVRTGHYECAEHLIACEADLNAKDREGDTPLHDAVRLNRYKMIRLLIMYGA
                     DLTIKNSAGKTPMDLVLNWQNGTKAIFDSLKENSYKTSRIATF"
BASE COUNT      623 a    391 c    479 g    447 t
ORIGIN

  Query Match             7.4%;  Score 154.4;  DB 4;  Length 1940;
  Best Local Similarity   84.9%;  Pred. No. 1.3e-24;
  Matches  185;  Conservative    0;  Mismatches   31;  Indels    2;  Gaps    1;

Qy       1859 CCTCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCT 1918
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db          1 CCTCTTGGGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGCTGGCT 60

Qy       1919 CCT--AATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGC 1976
               ||   | ||||||  ||  |||||||| ||||||||||||||||||||||||||||| ||
Db         61 ACTGGCAGGGGGGCAGGGCTGTTACTTGGGTTCCCAGGTTGGAAGATTATCTCACCCAGC 120

Qy       1977 CCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCT 2036
              || | ||||||||||  || | ||||||||||| ||||||||||||  |||||||||||| |
Db        121 CCAAACTATATAAGCCGAGCAGTGTGGAGGGGACCAGCAGGGCCATCTCCAGGGATTCAT 180

Qy       2037 TCCACGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
              |||||   |||||| || |||       || ||||||||
Db        181 TCCACAGGAGAAAACCACACAGACGCACTCCAGCCAAC 218


Search completed: November 18, 2003, 04:22:10
Job time : 7557.42 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:28:30 ; Search time 543.77 Seconds
(without alignments)
10295.963 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagttacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2071.6	99.9	2074	24	ABV73021	Human CARP protein
	2	367.8	17.7	2358	24	ABV73020	Mouse CARP protein
	3	322.6	15.6	2247	21	AAA10406	Murine cardiac ank
	4	229.6	11.1	1988	22	AAI93508	Human polynucleoti
	5	228	11.0	1901	22	AAH02910	Human shear stress
	6	228	11.0	1901	24	AAD27217	Human CAA58676 pro
	7	144.4	7.0	1889	24	AAS94903	Human DNA sequence
	8	144.4	7.0	1889	25	ABX77648	Differentially exp
	9	97.2	4.7	34488	22	AAF97854	Human neuroblastom
	10	97.2	4.7	121162	21	AAC66548	Human kinesin-like
	11	95.6	4.6	7141	22	AAK68280	Human immune/haema
c	12	95.4	4.6	6040	22	AAS28595	Genomic sequence #
c	13	94	4.5	4181	22	AAK84643	Human immune/haema
	14	94	4.5	7141	22	AAK68282	Human immune/haema
c	15	93.6	4.5	52845	22	AAK71437	Human immune/haema
c	16	93.4	4.5	29329	22	ABA18026	Human nervous syst
c	17	93.4	4.5	29329	22	ABA20511	Human nervous syst
c	18	93.4	4.5	29329	22	AAK70791	Human immune/haema
c	19	93.4	4.5	29329	22	AAK78512	Human immune/haema
	20	93.2	4.5	32220	22	AAL37568	Human musculoskele
	21	93.2	4.5	32220	25	ABX60556	cDNA encoding nove
	22	92.8	4.5	90220	24	ABK83576	Human cDNA differe
	23	92.8	4.5	99960	21	AAZ50905	Human TBC-1 partia
	24	92.6	4.5	19408	22	AAS42003	Genomic sequence #
	25	92.6	4.5	19408	22	AAK87230	Human immune/haema
	26	92.6	4.5	19408	22	AAK90644	Human digestive sy
	27	92.4	4.5	4678	22	ABA16110	Human nervous syst
	28	92.2	4.4	66804	24	ABK87050	Human transporter
	29	92	4.4	355	22	AAK60063	Human immune/haema
	30	91.4	4.4	3953	22	AAK85377	Human immune/haema
	31	91.4	4.4	3985	22	AAK85378	Human immune/haema
c	32	91.2	4.4	3073	24	ABQ77794	Human cGMP-inhibit
	33	91	4.4	26390	22	AAK65971	Human immune/haema
	34	91	4.4	56737	24	ABS69895	Human hypoxanthine
c	35	90.6	4.4	5248	22	AAK65595	Human immune/haema
c	36	90.6	4.4	5249	22	AAK65596	Human immune/haema
c	37	90.6	4.4	5249	22	AAK65597	Human immune/haema
c	38	90.6	4.4	118384	25	ABX56555	Human autoimmune d
	39	90.4	4.4	2778	21	AAC81712	Human secreted pro
c	40	90.4	4.4	19820	22	AAL36348	Human musculoskele
c	41	90.4	4.4	19820	25	ABX59336	cDNA encoding nove
	42	90.2	4.3	36785	22	AAK82208	Human immune/haema
	43	90	4.3	345	22	AAK76845	Human immune/haema
	44	90	4.3	222930	24	ABK84349	Human cDNA differe
	45	89.8	4.3	86592	25	ABZ22285	Human D-amino acid

ALIGNMENTS

RESULT 1
ABV73021
ID ABV73021 standard; DNA; 2074 BP.
XX
AC ABV73021;
XX
DT 08-JAN-2003 (first entry)
XX
DE Human CARP protein coding sequence upstream DNA fragment.
XX
KW Cardiac ankyrin repeat protein; CARP; cardiant; immunosuppressive;
KW antiinflammatory; gene therapy; antisense gene therapy; human; ds.
XX
OS Homo sapiens.
XX
PN WO200246220-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001WO-EP15412.
XX
PR 07-DEC-2000; 2000US-251582P.
XX
PA (AVET) AVENTIS PHARMA SA.
PA (REGC) UNIV CALIFORNIA.
PA (BENO/) BENOIT P.
XX
PI Schwartz B, Branellec D, Chien K;
XX
DR WPI; 2002-740642/80.
XX

PT New promoter sequence derived from a portion upstream of the coding
PT sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
PT the level and specificity of expression of a transgene in cardiac
PT muscle cells -
XX
PS Claim 5; Fig 2; 48pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising a fragment of a
CC sequence upstream of the coding part of the gene for the Cardiac Ankyrin
CC Repeat Protein (CARP). (I) is capable of inducing a specific expression
CC in vivo of a gene operably linked to (I), in cardiac cells. (I) or a
CC vector (IIb) comprising (I) is useful for the manufacture of a medicament
CC intended for the treatment of cardiac insufficiency, cardiac hypertrophy
CC and hypoxia, and for preventing rejection during cardiac transplant. An
CC expression cassette under the control of (I) is useful for encoding a
CC protein or RNA which is capable of activating the growth of cardiac
CC cells, reducing or suppressing an immune response, inducing angiogenesis,
CC correcting muscle contractility, cardiac hypertrophy, cardiac
CC insufficiency and myocarditis. (IIb) is useful for expressing a gene of
CC therapeutic interest in vivo, by isolating (IIb) and introducing (IIb) in
CC the cardiac tissue, under conditions so that the gene of interest is
CC expressed. (I), the vectors and the compositions are useful in clinical,
CC experimental, therapeutic and diagnostic fields, and in the treatment and
CC prevention of cardiac pathologies. (I) is also useful for generating
CC transgenic animals which constitute models for studying certain cardiac
CC pathologies. The transgenic animals are also useful for screening
CC molecules for their activity on the regulatory sequences of the gene
CC encoding the CARP protein. The present sequence represents the DNA
CC fragment upstream of the coding sequence of a human CARP protein.
XX
SQ Sequence 2074 BP; 612 A; 469 C; 416 G; 572 T; 5 other;

Query Match 99.9%; Score 2071.6; DB 24; Length 2074;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60

Qy         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120

Qy        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180

Qy        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240

Qy        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300

Qy        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360

Qy        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420

Qy        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480

Qy        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540

Qy        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600

Qy        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660

Qy        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720

Qy        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780

Qy        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTTTTCCTGAAACTATAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840

Qy        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900

Qy        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960

Qy        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020

Qy       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080

Qy       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140

Qy       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200

Qy       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260

Qy       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320

Qy       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380

Qy       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440

Qy       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500

Qy       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560

Qy       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620

Qy       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680

Qy       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740

Qy       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800

Qy       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860

Qy       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920

Qy       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980

Qy       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040

Qy       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||||||||||||
Db       2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
```

```
RESULT 2
ABV73020
ID   ABV73020 standard; DNA; 2358 BP.
XX
AC   ABV73020;
XX
DT   08-JAN-2003  (first entry)
XX
DE   Mouse CARP protein coding sequence upstream DNA fragment.
XX
KW   Cardiac ankyrin repeat protein; CARP; cardiant; immunosuppressive;
KW   antiinflammatory; gene therapy; antisense gene therapy; mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO200246220-A2.
XX
PD   13-JUN-2002.
XX
PF   05-DEC-2001; 2001WO-EP15412.
XX
PR   07-DEC-2000; 2000US-251582P.
XX
PA   (AVET ) AVENTIS PHARMA SA.
PA   (REGC ) UNIV CALIFORNIA.
PA   (BENO/) BENOIT P.
XX
PI   Schwartz B,  Branellec D,  Chien K;
XX
DR   WPI; 2002-740642/80.
XX
PT   New promoter sequence derived from a portion upstream of the coding
PT   sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
PT   the level and specificity of expression of a transgene in cardiac
PT   muscle cells  -
XX
PS   Claim 1; Fig 1; 48pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising a fragment of a
CC   sequence upstream of the coding part of the gene for the Cardiac Ankyrin
CC   Repeat Protein (CARP). (I) is capable of inducing a specific expression
CC   in vivo of a gene operably linked to (I), in cardiac cells. (I) or a
CC   vector (IIb) comprising (I) is useful for the manufacture of a medicament
CC   intended for the treatment of cardiac insufficiency, cardiac hypertrophy
CC   and hypoxia, and for preventing rejection during cardiac transplant. An
CC   expression cassette under the control of (I)  is useful for encoding a
CC   protein or RNA which is capable of activating the growth of cardiac
CC   cells, reducing or suppressing an immune response, inducing angiogenesis,
CC   correcting muscle contractility, cardiac hypertrophy, cardiac
CC   insufficiency and myocarditis. (IIb) is useful for expressing a gene of
CC   therapeutic interest in vivo, by isolating (IIb) and introducing (IIb) in
CC   the cardiac tissue, under conditions so that the gene of interest is
CC   expressed. (I), the vectors and the compositions are useful in clinical,
CC   experimental, therapeutic and diagnostic fields, and in the treatment and
CC   prevention of cardiac pathologies. (I) is also useful for generating
CC   transgenic animals which constitute models for studying certain cardiac
CC   pathologies. The transgenic animals are also useful for screening
CC   molecules for their activity on the regulatory sequences of the gene
CC   encoding the CARP protein. The present sequence represents the DNA
CC   fragment upstream of the coding sequence of a mouse CARP protein.
XX
SQ   Sequence 2358 BP; 636 A; 567 C; 542 G; 613 T; 0 other;
```

```
  Query Match             17.7%;  Score 367.8;  DB 24;  Length 2358;
  Best Local Similarity   76.4%;  Pred. No. 7.3e-76;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
             ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405
             |||||||||||| ||| |||  ||  |||   |||||||||||||| |||| |   |||||
Db       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657

Qy       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465
             |||||||||||||||||||||| |||  ||||| ||||| | ||||||||||  |||||
Db       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717

Qy       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524
             ||||||  |||||||||||||| ||||||||||||||||| |||||||| |     | |
Db       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777

Qy       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTT 1581
             |||  ||| | |||   |    ||  |||  | ||||||||||||    | |||  |
Db       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837

Qy       1582 TCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
              |  | || ||  |||| |||||||||   ||||||||||||||  |||   |||||||
Db       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897

Qy       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698
                || ||   | |||||||||||||| | |||||||||||| ||||||||||||  ||||
Db       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955

Qy       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757
              || |||| |||||||||||||| | ||||| |||||||||||||| | ||  ||  |  | ||
Db       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015

Qy       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817
              |||||| |||||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075

Qy       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877
              |||||||| ||||  | || |     |   |  ||  |||||||||  ||||||||||
Db       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130

Qy       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937
              ||| |||||||||||||||||||||||||||||| |||||  ||  ||||||| || |||
Db       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190

Qy       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997
              | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249
```

```
Qy       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056
              |||||||||||| |  ||||||||  ||||||| ||| | | || || ||||||||||
Db       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309

Qy       2057 A 2057
              |
Db       2310 A 2310
```

RESULT 3
AAA10406
ID AAA10406 standard; DNA; 2247 BP.
XX
AC AAA10406;
XX
DT 18-JUL-2000 (first entry)
XX
DE Murine cardiac ankyrin repeat protein (CARP) promoter.
XX
KW Cardiac ankyrin repeat protein; CARP promoter; murine; adenovirus vector;
KW cardiac specific; heart disease; gene therapy; ds.
XX
OS Mus musculus.
XX
PN WO200015821-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-US20730.
XX
PR 11-SEP-1998; 98US-0099960.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chien KR, Wang Y, Evans S;
XX
DR WPI; 2000-271457/23.
XX
PT Human type-5 recombinant adenovirus vector used for targeted gene
PT therapy for heart disease and evaluating gene function contains a
PT tissue-restricted promoter and inverted terminal repeat sequences -
XX
PS Claim 8; Page 29-30; 33pp; English.
XX
CC The invention relates to a human type-5 recombinant adenovirus vector
CC for achieving cardiac-restricted transcription of a gene of interest.
CC The vector comprises inverted terminal repeat (ITR) sequences from human
CC adeno-associated virus (AAV) type 2 (AAA10404-A10405) and a cardiac
CC tissue- specific promoter. In particular, the promoter is that of the
CC cardiomyocyte- restricted cardiac ankyrin repeat protein (CARP) gene.
CC The adenovirus vector is used for targetted gene therapy for heart
CC disease and for evaluating gene function. Cardiac restricted
CC transcription of a transgene in both neonatal and mature cardiac tissues
CC can be achieved to treat inherited and acquired heart diseases. The
CC vector is suitable for tissue-specific use in vivo and in vitro and
CC provides cardiac restricted transcription. The present sequence
CC represents the murine cardiac ankyrin repeat protein (CARP) promoter.
XX
SQ Sequence 2247 BP; 609 A; 549 C; 485 G; 589 T; 15 other;

  Query Match   15.6%; Score 322.6; DB 21; Length 2247;
  Best Local Similarity   76.3%; Pred. No. 2.5e-65;
  Matches 567; Conservative   0; Mismatches 154; Indels   22; Gaps   13;

```
Qy       1325 AAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCAT--GTGCTAATGGC 1382
              | ||  |||  ||||   | | | |||||||||||||| ||| |||  |   |  | |||||
Db       1499 AGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGTCACTAATTATGGC 1558

Qy       1383 CAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATGGT 1442
              ||||||||  ||||  |   |||||||||||||||||||||||||  |||  ||||| ||||
Db       1559 CAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGT 1618

Qy       1443 AAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACAGA 1502
              | | ||||||||||  |||||  ||||||          |||| ||||||||||||||||
Db       1619 ACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCATCCAGAGAATTCCAAACAGA 1678

Qy       1503 TATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATA 1562
              ||  ||||||||  |      | ||||  | ||||||||| ||||| ||| ||    ||||
Db       1679 TA-GGACAAGTGGCACCCAGACCCATCT-CCTTCCCCTCGGGCTGATTATCCCCAAAATA 1736

Qy       1563 GGATGTCC---TGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAA 1619
              ||||||||    |  |||  |  |  | || || |||| |||||||||   |||||||||
Db       1737 GGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAA 1796

Qy       1620 AGATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATAT 1678
              ||||||  |||   |||||||    || ||   | |||||||||||| | |||||||||
Db       1797 AGATATGGCTGTAAGTGTGATGCACAGTGC--TTGCATTTTCTTGATACGTTAGTCATAT 1854

Qy       1679 GAAAGCTGACAAAGAA--AAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTC 1736
              || ||||||||||||  ||||||  |||| |||||  |||||||  | |||||  ||||||||
Db       1855 GAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGT-GTGCAATATTAACAGGCAGCTGTC 1913

Qy       1737 CCCTGAC-TCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAG 1795
              |||||  |  ||  ||  |  | ||||||||  ||||||||||||| ||||  ||||  ||||||
Db       1914 CCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAG 1973

Qy       1796 GAATGGCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCC 1855
              ||||| ||||||||||||||||||||||| ||||  | ||  |       |   |  ||
Db       1974 GAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGG-----CCAGCTTGTCAT 2028

Qy       1856 CTCCCTCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTG 1915
              |||||||||  |||||||||||||| ||||||||||||||||||||||||||||||| |||
Db       2029 CTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTG 2088

Qy       1916 GCTCCTAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGG 1975
              ||  ||  |||| || || |||| |||| | |||||||| |||||||||||||||||| |
Db       2089 GCCACTGGTGGGAGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAG 2148

Qy       1976 CCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCC 2035
              ||| | |||||||| | | | |||||||||||||| |  ||||||||  ||||||| |||
Db       2149 CCCTA-CTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCA 2206

Qy       2036 TTC-CACGACAGAAAAACATACA 2057
              | | || || ||||||||||| |
Db       2207 TCCACAAGAGAGAAAAACATAGA 2229

RESULT 4
AAI93508
ID  AAI93508 standard; cDNA; 1988 BP.
XX
AC  AAI93508;
XX
DT  06-NOV-2001  (first entry)
XX
DE  Human polynucleotide SEQ ID NO 13568.
XX
KW  Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW  vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW  tissue growth factor; immunomodulatory; cancer; leukaemia;
KW  nervous system disorders; arthritis; inflammation; ss.
XX
OS  Homo sapiens.
XX
PN  WO200164835-A2.
XX
PD  07-SEP-2001.
XX
PF  26-FEB-2001; 2001WO-US04927.
XX
PR  28-FEB-2000; 2000US-0515126.
PR  18-MAY-2000; 2000US-0577409.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Tang YT,  Liu C,  Drmanac RT;
```

XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO13577.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 13568; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1988 BP; 632 A; 384 C; 472 G; 500 T; 0 other;

Query Match 11.1%; Score 229.6; DB 22; Length 1988;
Best Local Similarity 98.0%; Pred. No. 1.2e-43;
Matches 243; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
                | |  |||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db          2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
                ||||||||
Db        242 CAGCCAAC 249
```

RESULT 5
AAH02910
ID AAH02910 standard; DNA; 1901 BP.
XX
AC AAH02910;
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 73.
XX
KW Human; shear stress-response protein; vascular disease;
KW arteriosclerosis; ds.
XX
OS Homo sapiens.
XX
PN WO200125427-A1.
XX
PD 12-APR-2001.
XX
PF 02-OCT-2000; 2000WO-JP06840.
XX
PR 01-OCT-1999; 99JP-0280976.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (NOJI/) NOJIMA H.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI; 2001-266308/27.
DR P-PSDB; AAB90787.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 422-425; 678pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;

Query Match 11.0%; Score 228; DB 22; Length 1901;
Best Local Similarity 97.6%; Pred. No. 2.8e-43;
Matches 242; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
                | |  |||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db          2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
                ||||||||
Db        242 CAGCCAAC 249
```

RESULT 6
AAD27217
ID AAD27217 standard; DNA; 1901 BP.
XX
AC AAD27217;
XX
DT 09-APR-2002 (first entry)
XX
DE Human CAA58676 protein encoding EST clone X83703 DNA.
XX
KW Human; congestive heart failure; dilative cardiomyopathy; sudden death;
KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
KW heart muscle disease; conduction disorder; coronary heart disease;
KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
KW pulmonary heart disease; valvular heart disease; pericardial disease;
KW congenital heart disease; gene therapy; syncope; transgenic animal;
KW expressed sequence tag; EST; clone X83703; CAA58676 protein; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 250..1209
FT /*tag= a

```
FT                   /product= "Human CAA58676 protein"
FT   misc_feature    198..431
FT                   /*tag=  b
FT                   /note= "66268 cDNA fragment"
FT   misc_feature    1604..1753
FT                   /*tag=  b
FT                   /note= "S1MC01-1 cDNA fragment"
FT   misc_feature    1302..1306
FT                   /*tag=  b
FT                   /note= "AU-rich mRNA decay element"
FT   misc_feature    1391..1401
FT                   /*tag=  b
FT                   /note= "AU-rich mRNA decay element"
FT   misc_feature    1415..1423
FT                   /*tag=  b
FT                   /note= "AU-rich mRNA decay element"
XX
PN   WO200192567-A2.
XX
PD   06-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-EP06165.
XX
PR   30-MAY-2000; 2000US-207400P.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bunk D,  Reuner B,  Beck J,  Henkel T;
XX
DR   WPI; 2002-122073/16.
DR   P-PSDB; AAE16633.
XX
PT   Identifying a subject at risk for a heart disease e.g. congestive heart
PT   failure, dilative cardiomyopathy, heart muscle disease, by quantifying
PT   the polypeptide expressed by genes abnormally expressed in heart tissue
PT   -
XX
PS   Claim 2a; Fig 10b; 154pp; English.
XX
CC   The patent discloses novel target genes abnormally expressed in heart
CC   tissues and their corresponding proteins. The invention also relates to
CC   methods for assessing the expression level of these genes. The method
CC   is used for testing the predisposition of mammals and preferably humans
CC   for a heart disease or for an acute state of such a disease. It is also
CC   useful to treat diseases of the heart such as congestive heart failure,
CC   dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
CC   myopathy, specific heart muscle disease, rhythm and conduction disorders,
CC   syncope and sudden death, coronary heart disease, systemic arterial
CC   hypertension, pulmonary hypertension, pulmonary heart disease, valvular
CC   heart disease, congenital heart disease, pericardial disease and
CC   endocarditis. Sequences of the invention are also used in gene therapy.
CC   A transgenic non-human mammal comprising the sequences of the invention
CC   are useful for the development for medicaments for the treatments of
CC   heart diseases. The present DNA sequence is expressed sequence tag
CC   (EST) clone X83703 which encodes CAA58676 protein.
XX
SQ   Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;

  Query Match             11.0%;  Score 228;  DB 24;  Length 1901;
  Best Local Similarity   97.6%;  Pred. No. 2.8e-43;
  Matches  242;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1828 ATATTCAGCAGGGTTAGCTTGT-CCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
                | |  |||||||||||||||| |||||||||||||||||||||||||||||| |||
Db          2 AAAAACAGCAGGGTTAGCTTGTCCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCT 61

Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
                |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         62 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGG 121

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 181

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 2066
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTT 241

Qy       2067 CAGCCAAC 2074
                ||||||||
Db        242 CAGCCAAC 249


RESULT 7
AAS94903
ID   AAS94903 standard; DNA; 1889 BP.
XX
AC   AAS94903;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human DNA sequence #158 expressed during foam cell differentiation.
XX
KW   Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW   cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177389-A2.
XX
PD   18-OCT-2001.
XX
PF   04-APR-2001; 2001WO-US11128.
XX
PR   05-APR-2000; 2000US-195106P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Shiffman D,  Somogyi R,  Lawn R,  Seilhamer JJ,  Porter GJ,  Mikita T;
PI   Tai J;
XX
DR   WPI; 2002-010925/01.
XX
PT   Composition useful for diagnosis of conditions, disorders or diseases
PT   associated with atherosclerosis, comprises several polynucleotides that
PT   are differentially expressed in foam cell development -
XX
PS   Claim 1; Page 213; 315pp; English.
XX
CC   The present invention relates to the isolation of human polynucleotide
CC   sequences that are differentially expressed during foam cell
CC   differentiation. The polynucleotide sequences of the invention or a
CC   composition comprising these polynucleotides are useful as a high
CC   throughput method for detecting altered expression of one or more
CC   polynucleotides in a sample. The polynucleotides can be used in the
CC   diagnosis of disorders associated with foam cell development such as
CC   atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC   coronary artery disease. The polynucleotide sequences can also be used
CC   as PCR primers and probes. The polynucleotides of the invention are also
CC   useful in gene therapy. AAS94746-AAS95021 represent the human
CC   polynucleotide sequences of the invention which are differentially
CC   expressed during foam cell differentiation.
XX
SQ   Sequence 1889 BP; 613 A; 354 C; 452 G; 442 T; 28 other;

  Query Match             7.0%;  Score 144.4;  DB 24;  Length 1889;
  Best Local Similarity   99.3%;  Pred. No. 8.9e-24;
  Matches  145;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1929 GCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 1988
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 64

Qy       1989 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 2048
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 124
```

```
Qy       2049 AAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||
Db        125 AAACATACAAGACTCCTTCAGCCAAC 150
```

RESULT 8
ABX77648
ID ABX77648 standard; cDNA; 1889 BP.
XX
AC ABX77648;
XX
DT 09-APR-2003 (first entry)
XX
DE Differentially expressed breast cancer associated cDNA #143.
XX
KW Breast cancer; differential gene expression; BC-cDNA;
KW breast cancer diagnosis; breast cancer monitoring;
KW breast cancer treatment; breast cancer staging; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002156263-A1.
XX
PD 24-OCT-2002.
XX
PF 04-OCT-2001; 2001US-0974298.
XX
PR 05-OCT-2000; 2000US-238331P.
XX
PA (CHEN/) CHEN H.
XX
PI Chen H;
XX
DR WPI; 2003-182653/18.
XX
PT New cDNAs, which are differentially expressed in (metastatic) breast
PT cancer useful for diagnosing or staging, breast cancer, or for
PT monitoring the treatment of breast cancer in an individual -
XX
PS Claim 1; SEQ ID NO 182; 30pp; English.
XX
CC The invention describes a combination of cDNAs (designated BC-cDNAs),
CC which are differentially expressed in breast cancer. The combination
CC includes 152 cDNA sequences, or their complements. The protein encoded
CC by any of these BC-cDNAs is useful for screening several molecules or
CC compounds to identify at least one ligand that specifically binds the
CC protein, producing or preparing polyclonal or monoclonal antibodies, or
CC purifying antibodies from a sample. The antibodies, which specifically
CC bind the protein differentially expressed in breast cancer is useful for
CC detecting the expression of a protein in a sample. The BC-cDNAs are
CC also useful for diagnosing, monitoring the treatment of, or staging,
CC breast cancer. This sequence represents a differentially expressed
CC breast cancer associated cDNA.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=20020156263.
XX
SQ Sequence 1889 BP; 613 A; 354 C; 452 G; 469 T; 1 other;

Query Match 7.0%; Score 144.4; DB 25; Length 1889;
Best Local Similarity 99.3%; Pred. No. 8.9e-24;
Matches 145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1929 GCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 1988
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 64

Qy       1989 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 124
```

```
Qy       2049 AAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||
Db        125 AAACATACAAGACTCCTTCAGCCAAC 150
```

RESULT 9
AAF97854
ID AAF97854 standard; DNA; 34488 BP.
XX
AC AAF97854;
XX
DT 31-MAY-2001 (first entry)
XX
DE Human neuroblastoma cell line NB-1 1p36 nucleotide sequence SEQ ID NO:68.
XX
KW Human; chromosome 1; 1p36; neuroblastoma cell line; NB-1; anticancer;
KW tumour suppressor; human 1p36 homozygosity deletion domain; tumour;
KW diagnosis; ds.
XX
OS Homo sapiens.
XX
PN WO200116311-A1.
XX
PD 08-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-JP05930.
XX
PR 31-AUG-1999; 99JP-0245962.
PR 09-MAY-2000; 2000JP-0136266.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
PI Nakagawara A;
XX
DR WPI; 2001-226686/23.
XX
PT Human 1p36 homozygosity deletion domain from the 36-position of first
PT chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT in gene diagnosis of tumors as well as in developing anti-cancer drugs
PT -
XX
PS Example 8; Page 104-118; 226pp; Japanese.
XX
CC The present invention describes a homozygosity deletion domain
CC co-existing in the 36-position of the first chromosome short arm (1p36)
CC in human neuroblastoma. Also described are base sequences from the 1p36
CC position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC which are tumour suppressor genes in human neuroblastoma. The genes are
CC tumour suppressor genes, base sequence data of which are applicable as
CC tumour markers and reagents in studying mechanism of tumour body
CC formation, and gene diagnosis of tumours as well as in developing
CC anti-cancer drugs. AAF97787 to AAF97829 represent PCR primers used in
CC the exemplification of the present invention, and AAF97830 to AAF97874
CC represent sequences given in the exemplification of the present
CC invention.
XX
SQ Sequence 34488 BP; 9654 A; 6717 C; 6926 G; 11191 T; 0 other;

Query Match 4.7%; Score 97.2; DB 22; Length 34488;
Best Local Similarity 74.1%; Pred. No. 2.4e-12;
Matches 123; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Qy        258 AAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAAATAGGAGCTATACAAAG 317
              ||  | |||  ||  || | ||||| |||| |||||||||||||| ||| | ||||| ||
Db      20646 AATAACTAAGCACGGTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAG 20705

Qy        318 AAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGTGAAACATTCCATATATT 377
              |||||||||||| | ||| ||||| ||||||| |||||| ||||| | ||||||||  |
Db      20706 AAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTA 20765

Qy        378 AAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAGAAA 423
              ||||| ||| |||  | | | | |  |||| |  | |||  ||
```

```
Db      20766 AAAAAAAAAAAAAAGAAACAAAAACCACAAAAACAAAACAAAACAA 20811


RESULT 10
AAC66548
ID   AAC66548 standard; DNA; 121162 BP.
XX
AC   AAC66548;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.
XX
KW   Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;
KW   intracellular transport; neurological disorder; infertility;
KW   biallelic marker; spontaneous abortion; neonatal chromosome disorder;
KW   aneuploidy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200063375-A1.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-IB00562.
XX
PR   20-APR-1999;   99US-0130217.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Dufaure-Gare I,  Grel P;
XX
DR   WPI; 2000-665242/64.
XX
PT   An isolated or purified human kinesin-like protein (HKLP) encoding
PT   polynucleotide used to detect HKLP polynucleotides in a sample
PT   comprises a contiguous span of at least 12 nucleotides -
XX
PS   Claim 1; Page 143-175; 199pp; English.
XX
CC   The present invention describes the coding and protein sequences of the
CC   human kinesin-like protein HKLP. It is thought that the protein could be
CC   involved in neurological disorders, infertility, spontaneous abortion,
CC   neonatal chromosome disorders, aneuploidy and cancers. This is due to its
CC   function in the movement of microtubules. The protein shows homology to
CC   the murine KIF1A and KIF1B proteins. The sequences disclosed in the
CC   invention can be used in the isolation of similar human proteins and in
CC   vector production. In addition, the biallelic markers shown can be used
CC   in disease diagnosis and population studies.
XX
SQ   Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 21 other;

  Query Match             4.7%;  Score 97.2;  DB 21;  Length 121162;
  Best Local Similarity   74.1%;  Pred. No. 3.6e-12;
  Matches  123;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        258 AAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAAATAGGAGCTATACAAAG 317
              ||  | |||  ||  || | ||||| |||| |||||||||||||| ||| | ||||| ||
Db      47492 AATAACTAAGCACGGTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAG 47551

Qy        318 AAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGTGAAACATTCCATATATT 377
              ||||||||||||| | ||| ||||| ||||||| |||||| ||||| | |||||||| |
Db      47552 AAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTA 47611

Qy        378 AAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAGAAA 423
              ||||| ||| |||  | | | | |  |||| |  | |||  ||
Db      47612 AAAAAAAAAAAAAAGAAACAAAAACCACAAAAACAAAACAAAACAA 47657


RESULT 11
AAK68280
ID   AAK68280 standard; DNA; 7141 BP.
XX
AC   AAK68280;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23092.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
```

PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23092; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7141 BP; 2053 A; 1477 C; 1612 G; 1999 T; 0 other;

Query Match 4.6%; Score 95.6; DB 22; Length 7141;
Best Local Similarity 70.3%; Pred. No. 3.3e-12;
Matches 156; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

```
Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
             || | | |||||| | |||||||||||| |||  ||||||||||||||||||| |||||||
Db       3862 ATTTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGCCCAGGCTAGTTTTGAACT 3921

Qy       1196 CCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
             ||| |||| | ||||||| |||  ||| |||||||||||||| | ||||||||||||| ||
Db       3922 CCTGAGCTCAGGCAATCTGCCTACCTCGGCCTCCCAAAATGCTAGGATTACAGGTGTGAG 3981

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCAT 1314
              ||||  |||| ||| |  | ||  |||| ||  || |||||    |  |   || | |||
Db       3982 CCACCACACCTGGCCGTGAACT-TTTTCTTTCCCTTTATTTTCAAAAAAGAATGTTTCAT 4040

Qy       1315 TTTGAAGTAAAAATGCTCCAATTATTATGCTGTTTTAGAACA 1356
                | | |     || |   |||||  |  | || ||  | |
Db       4041 CATCCATTTTTTGTGATTAGATTATGGTTATATTATATTATA 4082
```

RESULT 12
AAS28595/c
ID AAS28595 standard; DNA; 6040 BP.
XX
AC AAS28595;
XX
DT 07-NOV-2001 (first entry)
XX
DE Genomic sequence #435 encoding for novel human respiratory antigen.
XX
KW Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;

KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
OS Homo sapiens.
XX
PN WO200155448-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01333.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476224/51.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
PS Disclosure; SED ID No 1029; 546pp; English.
XX
CC The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6040 BP; 1446 A; 1488 C; 1420 G; 1686 T; 0 other;

Query Match 4.6%; Score 95.4; DB 22; Length 6040;
Best Local Similarity 72.8%; Pred. No. 3.5e-12;
Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
             || | ||   ||| | || |||||||| ||| ||||| ||| ||||||| || |||||||
Db       3417 ATTTTTATATTTTTTAGTGGAGATGGGGTTTCGCCATATTGGCCAGGCTGGTCTTGAACT 3358

Qy       1196 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 1255
             ||| | || ||| |||| |||| ||| ||||| |||  || ||||||||||||| || |||
Db       3357 CCTGACCTCAGCGATCTGCCTGCCTCAGCCTCTCAATGTGCTGGGATTACAGGCGTGAGC 3298

Qy       1256 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCAC 1304
             ||||||||| ||| | |   |  ||   |||  | | ||| |||   ||
Db       3297 CACTGCACCTGGCCGTTTTTTTTTTCTTTTTCTTTTTTTTTTTGGAGAC 3249
```

RESULT 13
AAK84643/c
ID AAK84643 standard; DNA; 4181 BP.
XX
AC AAK84643;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39455.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

```
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 39455; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 4181 BP; 1137 A; 1031 C; 985 G; 1028 T; 0 other;

  Query Match             4.5%;  Score 94;  DB 22;  Length 4181;
  Best Local Similarity   58.9%;  Pred. No. 6.6e-12;
  Matches  178;  Conservative    1;  Mismatches  122;  Indels    1;  Gaps    1;

Qy       1022 TTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAAT 1081
             || || |   || | ||    | ||||   | |   |||      ||||||||
Db       4068 TTTTGGCTCACTGCAACCTCCGCCTCCTGGGCTTAAGCAATTCTCCTGCTTCAGCTTCCC 4009

Qy       1082 GACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATAT 1141
             |||  | | ||| | | | ||       |          |   |||:|      | | | |
Db       4008 GACTAACTGAGATTACAGGCACCCGCCACCATGCCCAGCTAATTTTTTTGTATTTTTTTT 3949

Qy       1142 AGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAAG 1201
               | ||| | || |||| ||| |||  |||||||| ||||||| || |||||  ||| |
Db       3948 TTTTTTTTTAGTTGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAATGCCTGAT 3889

Qy       1202 CT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260
             || | |  |||  || | ||| ||||||||||| ||||||||||||||| || ||||||||
Db       3888 CTCAGGTGATCCACCCGCCTCGGCCTCCCAAAGTGTTGGGATTACAGGCGTGAGCCACTG 3829

Qy       1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
             | ||||||  |  ||  ||||| |||| | | ||      || ||   ||   ||||  |
Db       3828 CGCCCGGCCCAGTGCCAGTTTCTTTTAGTGTTTTAGGGAACAACCTAGACAGATTTATCA 3769

Qy       1321 GT 1322
             ||
Db       3768 GT 3767


RESULT 14
AAK68282
ID   AAK68282 standard; DNA; 7141 BP.
XX
AC   AAK68282;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23094.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
```

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23094; 3071pp + Sequence Listing; English.
XX

```
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 7141 BP; 2053 A; 1476 C; 1615 G; 1997 T; 0 other;

  Query Match             4.5%;  Score 94;  DB 22;  Length 7141;
  Best Local Similarity   69.8%;  Pred. No. 7.9e-12;
  Matches  155;  Conservative    0;  Mismatches   65;  Indels    2;  Gaps    2;

Qy      1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
             || | | |||||| | |||||||||| |||  ||||||||||||||||||| |||||||
Db      3862 ATTTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGCCCAGGCTAGTTTTGAACT 3921

Qy      1196 CCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
             ||| |||| | ||||||| |||  ||| ||||||||||||| | ||||||||||||| ||
Db      3922 CCTGAGCTCAGGCAATCTGCCTACCTCGGCCTCCCAAAATGCTAGGATTACAGGTGTGAG 3981

Qy      1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCAT 1314
             ||||  |||| ||| |  |  |  |||| ||  || |||||    |  |   || | |||
Db      3982 CCACCACACCTGGCCGTGAAAT-TTTTCTTTCCCTTTATTTTCAAAAAAGAATGTTTCAT 4040

Qy      1315 TTTGAAGTAAAAAATGCTCCAATTATTATGCTGTTTTAGAACA 1356
               |  | |      || |   |||||  |  | || ||  | |
Db      4041 CATCCATTTTTTGTGATTAGATTATGGTTATATTATATTATA 4082


RESULT 15
AAK71437/c
ID   AAK71437 standard; DNA; 52845 BP.
XX
AC   AAK71437;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26249.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
```

PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 26249; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 52845 BP; 14698 A; 11081 C; 10505 G; 16561 T; 0 other;

Query Match 4.5%; Score 93.6; DB 22; Length 52845;
Best Local Similarity 67.3%; Pred. No. 1.9e-11;
Matches 132; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | |||||| ||||| ||||||| |||  |||| ||| ||||||| || |||||
Db      51170 TAATTTTTGTATTTTTGGTAAAGATGGGGTTTCACCATTTTGGCCAGGCTGGTCTTGAAT 51111

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
             |||| | | |||  |||| |||| |||||||||||||| || |||||||||||||| |
Db      51110 TCCTGACCCAAGTGATCTCCCTGCCTCTGCCTCCCAAAGTGCTGGGATTACAGGTGTGAA 51051

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCAT 1314
              |||| ||||| ||   ||| |   |  |  |||  | | |   |  ||| |
Db      51050 TCACTACACCCAGCCTTTAGATTTGTGTAAGTACATTCTATGAGGTTCACACAATGATGA 50991

Qy       1315 TTTGAAGTAAAAATGC 1330
                |.   ||| |||||
Db      50990 AATTGCTTAAGAATGC 50975
```

Search completed: November 17, 2003, 23:52:55
Job time : 550.27 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:50 ; Search time 120.734 Seconds
(without alignments)
7582.209 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagttacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	92.2	4.4	66804	4	US-09-740-041-3	Sequence 3, Appli
	2	89.4	4.3	6669	3	US-09-212-971-5	Sequence 5, Appli
	3	89.4	4.3	6669	3	US-08-800-929A-5	Sequence 5, Appli
	4	89.4	4.3	6669	4	US-09-617-053A-5	Sequence 5, Appli
	5	89.4	4.3	75395	4	US-09-984-890-3	Sequence 3, Appli
	6	87.8	4.2	59065	4	US-09-813-817-3	Sequence 3, Appli
	7	87.8	4.2	59065	4	US-09-978-197-3	Sequence 3, Appli
	8	86.8	4.2	956	4	US-09-641-638-36	Sequence 36, Appl
	9	86.8	4.2	20674	4	US-09-641-638-651	Sequence 651, App
	10	86.2	4.2	99500	4	US-09-798-096-10	Sequence 10, Appl
c	11	86	4.1	29629	4	US-09-729-995-3	Sequence 3, Appli
c	12	85.8	4.1	72604	4	US-09-268-992-7	Sequence 7, Appli
c	13	85.8	4.1	72604	4	US-09-657-474-7	Sequence 7, Appli
c	14	85.8	4.1	75395	4	US-09-984-890-3	Sequence 3, Appli
c	15	85.2	4.1	90541	4	US-09-759-359A-3	Sequence 3, Appli
c	16	85	4.1	330	3	US-09-078-294-24	Sequence 24, Appl
	17	84.6	4.1	4550	3	US-09-103-663-35	Sequence 35, Appl
	18	84.4	4.1	72604	4	US-09-268-992-7	Sequence 7, Appli
	19	84.4	4.1	72604	4	US-09-657-474-7	Sequence 7, Appli
	20	84.2	4.1	1001	4	US-09-641-638-199	Sequence 199, App
	21	84.2	4.1	1001	4	US-09-641-638-200	Sequence 200, App
c	22	84.2	4.1	3001	4	US-09-539-333D-142	Sequence 142, App
c	23	84.2	4.1	35060	3	US-08-814-095-7	Sequence 7, Appli
	24	84.2	4.1	46718	4	US-09-816-093-3	Sequence 3, Appli
c	25	84	4.1	53332	4	US-09-801-861-3	Sequence 3, Appli
c	26	84	4.1	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	27	84	4.1	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	28	84	4.1	246240	2	US-08-724-394A-22	Sequence 22, Appl
	29	83.8	4.0	43950	4	US-09-735-934A-3	Sequence 3, Appli
	30	83.8	4.0	43950	4	US-10-060-332-3	Sequence 3, Appli
c	31	83.8	4.0	84495	4	US-09-797-906-3	Sequence 3, Appli
c	32	83.8	4.0	162450	4	US-09-345-882-1	Sequence 1, Appli
	33	83.6	4.0	63588	4	US-09-873-404-3	Sequence 3, Appli
c	34	83.4	4.0	66804	4	US-09-740-041-3	Sequence 3, Appli
	35	83.2	4.0	4741	1	US-07-695-472B-4	Sequence 4, Appli
	36	83.2	4.0	4741	4	US-09-106-375-4	Sequence 4, Appli
	37	83.2	4.0	4742	1	US-08-250-740-35	Sequence 35, Appl
c	38	83.2	4.0	55298	4	US-09-491-356C-1	Sequence 1, Appli
	39	83	4.0	98844	4	US-09-791-211-10	Sequence 10, Appl
c	40	83	4.0	116592	4	US-09-818-512-3	Sequence 3, Appli
c	41	82.6	4.0	152331	3	US-09-128-155-16	Sequence 16, Appl
	42	82.4	4.0	4545	4	US-09-569-852B-5	Sequence 5, Appli
	43	82.4	4.0	20966	4	US-09-569-852B-1	Sequence 1, Appli
c	44	82.4	4.0	65042	4	US-09-784-316-3	Sequence 3, Appli
c	45	82.2	4.0	8982	3	US-08-976-255-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-740-041-3
; Sequence 3, Application US/09740041
; Patent No. 6562593
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001001
; CURRENT APPLICATION NUMBER: US/09/740,041
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 66804
; TYPE: DNA
; ORGANISM: Human
US-09-740-041-3

Query Match 4.4%; Score 92.2; DB 4; Length 66804;
Best Local Similarity 73.3%; Pred. No. 3.4e-13;
Matches 118; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | | |||||| |||||||||| |||||||||||||||| ||||||| || | ||||
Db      40127 TAATTTTTGTATTTTTGGTAGAGATAGGATTTTGCCATGTTGGCCAGGCTGGTCTCGAAC 40186

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              |||| |  || |  |||  ||   ||| |||||||||| |||||||||||| |  ||||
Db      40187 TCCTGACTTAGGTGATCCGCCCACCTCGGCCTCCCAAACTGTTGGGATTACAAGCATAAG 40246

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTT 1295
              ||||||  || |||   |   |  ||||| |  | || | |
Db      40247 CCACTGTGCCTGGCCTTTTTTTTTTTTCAGTCTCGCTCTGT 40287
```

RESULT 2
US-09-212-971-5
; Sequence 5, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: MacKenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE

```
;  TITLE OF INVENTION: DISEASE
;  FILE REFERENCE: 07891/009002
;  CURRENT APPLICATION NUMBER: US/09/212,971B
;  CURRENT FILING DATE: 1998-12-16
;  EARLIER APPLICATION NUMBER: 60/017,354
;  EARLIER FILING DATE: 1996-04-26
;  EARLIER APPLICATION NUMBER: 60/030,590
;  EARLIER FILING DATE: 1996-11-14
;  EARLIER APPLICATION NUMBER: 08/800,929
;  EARLIER FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 6669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (3677)...(3951)
;   OTHER INFORMATION: n can be any nucleotide
US-09-212-971-5

  Query Match             4.3%;  Score 89.4;  DB 3;  Length 6669;
  Best Local Similarity   73.1%;  Pred. No. 7.2e-13;
  Matches  128;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | ||||||  | ||||||||||| |||  |||| ||| ||||||| || ||||||
Db       1230 TAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATTTTGGCCAGGCTGGTTTTGAAC 1289

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
             |||| | || |||  |||  | |||||  |||||||||||| |||||||||||  || |
Db       1290 TCCTGACCTCAAGTGATCCACTTGTCTTGGCCTCCCAAAATGCTGGGATTACAGGCGTGA 1349

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTG 1308
             ||||||||||| ||| ||   ||   ||  ||  |||| |||     | ||  ||
Db       1350 GCCACTGCACCAGGCAGAGGCCTCTGTTTTTTATCTCTTTTTGGCCTCTACAGTG 1404


RESULT 3
US-08-800-929A-5
; Sequence 5, Application US/08800929A
; Patent No. 6133437
;  GENERAL INFORMATION:
;    APPLICANT:  Korneluk, Robert G
;    APPLICANT:  MacKenzie, Alexander E
;    APPLICANT:  Liston, Peter
;    APPLICANT:  Baird, Stephen
;    APPLICANT:  Tsang, Benjamin K
;    APPLICANT:  Pratt, Christine
;    TITLE OF INVENTION:  DETECTION AND MODULATION OF
;    TITLE OF INVENTION:  IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;    TITLE OF INVENTION:  DISEASE
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,929A
;      FILING DATE:  13-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/030,590
;      FILING DATE:  14-NOV-1996
;      APPLICATION NUMBER:  60/017,354
;      FILING DATE:  26-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  07891/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-428-0200
;      TELEFAX:  617-428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-800-929A-5

  Query Match             4.3%;  Score 89.4;  DB 3;  Length 6669;
  Best Local Similarity   73.1%;  Pred. No. 7.2e-13;
  Matches  128;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | ||||||  | ||||||||||| |||  |||| ||| ||||||| || ||||||
Db       1230 TAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATTTTGGCCAGGCTGGTTTTGAAC 1289

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
             |||| | || |||  |||  | |||||  |||||||||||| |||||||||||  || |
Db       1290 TCCTGACCTCAAGTGATCCACTTGTCTTGGCCTCCCAAAATGCTGGGATTACAGGCGTGA 1349

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTG 1308
             ||||||||||| ||| ||   ||   ||  ||  |||| |||     | ||  ||
Db       1350 GCCACTGCACCAGGCAGAGGCCTCTGTTTTTTATCTCTTTTTGGCCTCTACAGTG 1404


RESULT 4
US-09-617-053A-5
; Sequence 5, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
;  APPLICANT: Korneluk, Robert G
;  APPLICANT:  MacKenzie, Alexander E
;  APPLICANT:  Liston, Peter
;  APPLICANT:  Baird, Stephen
;  APPLICANT:  Tsang, Benjamin K
;  APPLICANT:  Pratt, Christine
;  TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
;  TITLE OF INVENTION:  NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;  TITLE OF INVENTION:  DISEASE
;  FILE REFERENCE: 07891/009003
;  CURRENT APPLICATION NUMBER: US/09/617,053A
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/800,929
;  PRIOR FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 6669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (3677)...(3951)
;   OTHER INFORMATION: n can be any nucleotide
US-09-617-053A-5

  Query Match             4.3%;  Score 89.4;  DB 4;  Length 6669;
  Best Local Similarity   73.1%;  Pred. No. 7.2e-13;
  Matches  128;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | ||||||  | ||||||||||| |||  |||| ||| ||||||| || ||||||
```

```
Db       1230 TAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATTTTGGCCAGGCTGGTTTTGAAC 1289

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
              |||| | || |||  |||  | |||||  |||||||||||| |||||||||||| || |
Db       1290 TCCTGACCTCAAGTGATCCACTTGTCTTGGCCTCCCAAAATGCTGGGATTACAGGCGTGA 1349

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTG 1308
              |||||||||||| ||| ||   ||   ||  |||| |||     | ||  ||
Db       1350 GCCACTGCACCAGGCAGAGGCCTCTGTTTTTTATCTCTTTTTGGCCTCTACAGTG 1404


RESULT 5
US-09-984-890-3
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001306
;  CURRENT APPLICATION NUMBER: US/09/984,890
;  CURRENT FILING DATE:  2001-10-31
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 75395
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(75395)
;   OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

  Query Match             4.3%;  Score 89.4;  DB 4;  Length 75395;
  Best Local Similarity   72.1%;  Pred. No. 1.8e-12;
  Matches  129;  Conservative    1;  Mismatches   48;  Indels    1;  Gaps    1;

Qy       1090 CAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATATAGTATTTG 1149
              |||   |||||||       |  ||  | |  |||||:|  | | | | | | |||||||
Db      13177 CAGGCACCTGCCACCATACCTGGCTAATTTTTTTTTTTTTTTTTTTTTTTTGTATTTT 13236

Qy       1150 TGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAAGCTAAGCAA 1209
              | |||||||||||| |||  |||||||| ||||||| || |||||||||| | || |   |
Db      13237 TAGTAGAGATGGGGTTTCACCATGTTGACCAGGCTGGTCTTGAACTCCTGACCTCA-TGA 13295

Qy       1210 TCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 1268
              |||  ||   ||| |||||||||| || |||||||||||| || ||||||||||||| |
Db      13296 TCTGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCACCCGCC 13354


RESULT 6
US-09-813-817-3
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001178
;  CURRENT APPLICATION NUMBER: US/09/813,817
;  CURRENT FILING DATE:  2001-03-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 59065
;   TYPE: DNA
;   ORGANISM: Human
US-09-813-817-3

  Query Match             4.2%;  Score 87.8;  DB 4;  Length 59065;
  Best Local Similarity   78.1%;  Pred. No. 4.2e-12;
  Matches  118;  Conservative    0;  Mismatches   32;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | |  ||||| | |||||||||||| |||||||||||| ||| |||||| | ||||
Db      11804 TAATTTTTATATTTTTAGTAGAGATGGGGTTTTGCCATGTTGGCCAAGCTAGTCTCGAAC 11863

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
              |||| | || |||  |||  |||| ||| |||||||||| || |||||||||||||| |
Db      11864 TCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGA 11923

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCAT 1284
              |||||||||||  |||||  ||  |   || |
Db      11924 GCCACTGCACCTGGCTGGAAGGAGTGATCTT 11954


RESULT 7
US-09-978-197-3
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001178DIV
;  CURRENT APPLICATION NUMBER: US/09/978,197
;  CURRENT FILING DATE:  2001-10-17
;  PRIOR APPLICATION NUMBER: 09/813,817
;  PRIOR FILING DATE: 2001-03-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 59065
;   TYPE: DNA
;   ORGANISM: Human
US-09-978-197-3

  Query Match             4.2%;  Score 87.8;  DB 4;  Length 59065;
  Best Local Similarity   78.1%;  Pred. No. 4.2e-12;
  Matches  118;  Conservative    0;  Mismatches   32;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | |  ||||| | |||||||||||| |||||||||||| ||| |||||| | ||||
Db      11804 TAATTTTTATATTTTTAGTAGAGATGGGGTTTTGCCATGTTGGCCAAGCTAGTCTCGAAC 11863

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
              |||| | || |||  |||  |||| ||| |||||||||| || |||||||||||||| |
Db      11864 TCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGA 11923

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCAT 1284
              |||||||||||  |||||  ||  |   || |
Db      11924 GCCACTGCACCTGGCTGGAAGGAGTGATCTT 11954


RESULT 8
US-09-641-638-36
; Sequence 36, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Cohen, Annick
;  TITLE OF INVENTION:  BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;  TITLE OF INVENTION:  GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;  FILE REFERENCE:  GENSET.051CP1
;  CURRENT APPLICATION NUMBER: US/09/641,638
;  CURRENT FILING DATE:  2000-08-16
;  PRIOR APPLICATION NUMBER:  US 09/502,330
;  PRIOR FILING DATE:  2000-02-11
;  PRIOR APPLICATION NUMBER:  US 60/133,200
;  PRIOR FILING DATE:  1999-05-07
```

```
;  PRIOR APPLICATION NUMBER:  US 09/275,267
;  PRIOR FILING DATE:  1999-03-23
;  PRIOR APPLICATION NUMBER:  US 60/119,917
;  PRIOR FILING DATE:  1999-02-12
;  NUMBER OF SEQ ID NOS:  1304
;  SOFTWARE:  Patent.pm
; SEQ ID NO 36
;   LENGTH: 956
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 478
;   OTHER INFORMATION: 10-349-47 : polymorphic base C or T
;   NAME/KEY: misc_binding
;   LOCATION: 458..477
;   OTHER INFORMATION: 10-349-47.mis1, potential
;   NAME/KEY: misc_binding
;   LOCATION: 479..498
;   OTHER INFORMATION: 10-349-47.mis2, potential complement
;   NAME/KEY: primer_bind
;   LOCATION: 432..451
;   OTHER INFORMATION: upstream amplification primer
;   NAME/KEY: primer_bind
;   LOCATION: 829..848
;   OTHER INFORMATION: downstream amplification primer, complement
;   NAME/KEY: misc_binding
;   LOCATION: 466..490
;   OTHER INFORMATION: 10-349-47 potential probe
US-09-641-638-36

  Query Match             4.2%;  Score 86.8;  DB 4;  Length 956;
  Best Local Similarity   76.6%;  Pred. No. 1.5e-12;
  Matches  118;  Conservative    1;  Mismatches   34;  Indels    1;  Gaps    1;

Qy       1116 GTCTCATTTTTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTT 1175
              | | |  | | : | |   ||  | | |||||| | |||||||||||| ||| ||||| ||
Db          9 GCCACCATGTGGGCCTGGCTAATTTTTGTATTTTTTGTAGAGATGGGGTTTCGCCATATT 68

Qy       1176 GCCCAGGCTAGTATTGAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAAT 1234
              ||||||||| || ||||||| ||  ||| ||| ||||| |||| ||| ||||||||||| |
Db         69 GCCCAGGCTGGTCTTGAACTTCTGGGCTCAAGTAATCTGCCTGCCTCGGCCTCCCAAAGT 128

Qy       1235 GTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 1268
              | |||||||||||| ||| ||||||||||||| |||
Db        129 GCTGGGATTACAGCTGTGAGCCACTGCACCTGGC 162


RESULT 9
US-09-641-638-651
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Cohen, Annick
;  TITLE OF INVENTION:  BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;  TITLE OF INVENTION:  GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;  FILE REFERENCE:  GENSET.051CP1
;  CURRENT APPLICATION NUMBER: US/09/641,638
;  CURRENT FILING DATE:  2000-08-16
;  PRIOR APPLICATION NUMBER:  US 09/502,330
;  PRIOR FILING DATE:  2000-02-11
;  PRIOR APPLICATION NUMBER:  US 60/133,200
;  PRIOR FILING DATE:  1999-05-07
;  PRIOR APPLICATION NUMBER:  US 09/275,267
;  PRIOR FILING DATE:  1999-03-23
;  PRIOR APPLICATION NUMBER:  US 60/119,917
;  PRIOR FILING DATE:  1999-02-12
;  NUMBER OF SEQ ID NOS:  1304
;  SOFTWARE:  Patent.pm
; SEQ ID NO 651
;   LENGTH: 20674
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1123..3123
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 3124..3297
;   OTHER INFORMATION: exon 1
;   NAME/KEY: exon
;   LOCATION: 3871..4072
;   OTHER INFORMATION: exon 2
;   NAME/KEY: exon
;   LOCATION: 5552..5633
;   OTHER INFORMATION: exon 3
;   NAME/KEY: exon
;   LOCATION: 5758..5880
;   OTHER INFORMATION: exon 4
;   NAME/KEY: exon
;   LOCATION: 5996..6099
;   OTHER INFORMATION: exon 5
;   NAME/KEY: exon
;   LOCATION: 6349..6509
;   OTHER INFORMATION: exon 6
;   NAME/KEY: exon
;   LOCATION: 7379..7522
;   OTHER INFORMATION: exon 7
;   NAME/KEY: exon
;   LOCATION: 8645..8854
;   OTHER INFORMATION: exon 8
;   NAME/KEY: exon
;   LOCATION: 12254..12340
;   OTHER INFORMATION: exon 9
;   NAME/KEY: exon
;   LOCATION: 12854..13023
;   OTHER INFORMATION: exon 10
;   NAME/KEY: exon
;   LOCATION: 13308..13429
;   OTHER INFORMATION: exon 11
;   NAME/KEY: exon
;   LOCATION: 16567..16667
;   OTHER INFORMATION: exon 12
;   NAME/KEY: exon
;   LOCATION: 16775..16945
;   OTHER INFORMATION: exon 13
;   NAME/KEY: exon
;   LOCATION: 17063..17554
;   OTHER INFORMATION: exon 14
;   NAME/KEY: misc_feature
;   LOCATION: 17555..20674
;   OTHER INFORMATION: 3'regulatory region
;   NAME/KEY: allele
;   LOCATION: 1128
;   OTHER INFORMATION: 10-508-191  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 1182
;   OTHER INFORMATION: 10-508-245  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 1559
;   OTHER INFORMATION: 10-509-284  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 1570
;   OTHER INFORMATION: 10-509-295  : deletion of C
;   NAME/KEY: allele
;   LOCATION: 1827
;   OTHER INFORMATION: 10-510-173  : variable motif ATTTA or TTTTTT
;   NAME/KEY: allele
;   LOCATION: 2048
;   OTHER INFORMATION: 10-511-62  : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 2323
;   OTHER INFORMATION: 10-511-337  : insertion of T
```

; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 5903
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6019
; OTHER INFORMATION: 10-346-141 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6141
; OTHER INFORMATION: 10-346-263 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6183
; OTHER INFORMATION: 10-346-305 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6338
; OTHER INFORMATION: 10-347-74 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6375
; OTHER INFORMATION: 10-347-111 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6429
; OTHER INFORMATION: 10-347-165 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6467
; OTHER INFORMATION: 10-347-203 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6484
; OTHER INFORMATION: 10-347-220 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6534
; OTHER INFORMATION: 10-347-271 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 6611
; OTHER INFORMATION: 10-347-348 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 7668
; OTHER INFORMATION: 10-348-391 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8608
; OTHER INFORMATION: 10-349-47 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 8658
; OTHER INFORMATION: 10-349-97 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8703
; OTHER INFORMATION: 10-349-142 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 8777
; OTHER INFORMATION: 10-349-216 : deletion of CTG
; NAME/KEY: allele
; LOCATION: 8785
; OTHER INFORMATION: 10-349-224 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 8926
; OTHER INFORMATION: 10-349-368 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12171
; OTHER INFORMATION: 10-350-72 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12429
; OTHER INFORMATION: 10-350-332 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13341
; OTHER INFORMATION: 10-507-170 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 13492
; OTHER INFORMATION: 10-507-321 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 13524
; OTHER INFORMATION: 10-507-353 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13535

Query Match 4.2%; Score 86.8; DB 4; Length 20674;
Best Local Similarity 76.6%; Pred. No. 5e-12;
Matches 118; Conservative 1; Mismatches 34; Indels 1; Gaps 1;

```
Qy       1116 GTCTCATTTTTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTT 1175
              | | |  | |.: | |   ||  | | |||||| | |||||||||||| ||| ||||| ||
Db       8139 GCCACCATGTGGGCCTGGCTAATTTTTGTATTTTTTGTAGAGATGGGGTTTCGCCATATT 8198

Qy       1176 GCCCAGGCTAGTATTGAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAAT 1234
              ||||||||| || ||||||| ||  ||| ||| ||||| |||| ||| ||||||||||
Db       8199 GCCCAGGCTGGTCTTGAACTTCTGGGCTCAAGTAATCTGCCTGCCTCGGCCTCCCAAAGT 8258

Qy       1235 GTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 1268
              | |||||||||||| ||| |||||||||||| |||
Db       8259 GCTGGGATTACAGCTGTGAGCCACTGCACCTGGC 8292
```

RESULT 10
US-09-798-096-10
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

Query Match 4.2%; Score 86.2; DB 4; Length 99500;
Best Local Similarity 75.5%; Pred. No. 1.3e-11;
Matches 120; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```
Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
```

```
              || | | || ||| | |||||| |||| ||| |||||||| ||||||| || | |||||
Db      32693 ATTTTTTGTGTTTTTATTAGAGGTGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACT 32752

Qy       1196 CCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              ||| | || |||||| ||  |||| ||| |||||||||| || ||||||||||||  | ||
Db      32753 CCTGACCTCAAGCATTCCACCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAG 32812

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTAT 1293
              |||||||||||||   |  |   |||||||||  || |
Db      32813 CCACTGCACCCGGCCTCTTCCCTTTTTCATTTATGCTGT 32851


RESULT 11
US-09-729-995-3/c
; Sequence 3, Application US/09729995
; Patent No. 6426206
; GENERAL INFORMATION:
;  APPLICANT: WEI, Ming-Hui et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL000904
;  CURRENT APPLICATION NUMBER: US/09/729,995
;  CURRENT FILING DATE:  2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 29629
;   TYPE: DNA
;   ORGANISM: Human
US-09-729-995-3

  Query Match             4.1%;  Score 86;  DB 4;  Length 29629;
  Best Local Similarity   77.6%;  Pred. No. 9.2e-12;
  Matches  104;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | | |||||| | |||||||||| |||  |||||||| ||||||| || |  |||
Db      16878 TAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCTAAC 16819

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              ||||||   | |  |||| ||   ||| |||||||||| || ||||||||||| || ||
Db      16818 TCCTAACTCAGGTGATCTGCCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAG 16759

Qy       1255 CCACTGCACCCGGC 1268
              |||| |||||||||
Db      16758 CCACCGCACCCGGC 16745


RESULT 12
US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT: Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/268,992
;  CURRENT FILING DATE: 1999-03-16
;  EARLIER APPLICATION NUMBER: 09/236,134
;  EARLIER FILING DATE: 1999-01-22
;  EARLIER APPLICATION NUMBER: 60/106,056
;  EARLIER FILING DATE: 1998-10-28
;  EARLIER APPLICATION NUMBER: 60/088,312
;  EARLIER FILING DATE: 1998-06-05
;  EARLIER APPLICATION NUMBER: 60/078,044
;  EARLIER FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

  Query Match             4.1%;  Score 85.8;  DB 4;  Length 72604;
  Best Local Similarity   72.5%;  Pred. No. 1.5e-11;
  Matches  111;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              | | | | |||||| |  |||||||||| |||  ||||||||  |||||| || | ||||
Db      57580 TTTTTTTTGTATTTTTAATAGAGATGGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAAC 57521

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              |||| | ||   | |||  |||| ||  |||||||||||| ||||||||||||  | ||
Db      57520 TCCTGACCTCGTCGATCCGCCTGCCTTGGCCTCCCAAAATGCTGGGATTACAGGCATGAG 57461

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTA 1287
              ||||  ||||||||||   |  |  |   ||||
Db      57460 CCACCACACCCGGCTGTATGAAGCATGTTTTTA 57428


RESULT 13
US-09-657-474-7/c
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
;  APPLICANT: Chen, H.
;  APPLICANT:  Freimer, N.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CHROMOSOME-18p RELATED DISORDERS
;  FILE REFERENCE: 7853-138
;  CURRENT APPLICATION NUMBER: US/09/657,474
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: 09/268,992
;  PRIOR FILING DATE: 1999-03-16
;  PRIOR APPLICATION NUMBER: 09/236,134
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/106,056
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/088,312
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/078,044
;  PRIOR FILING DATE: 1998-03-16
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

  Query Match             4.1%;  Score 85.8;  DB 4;  Length 72604;
  Best Local Similarity   72.5%;  Pred. No. 1.5e-11;
  Matches  111;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              | | | | |||||| |  |||||||||| |||  ||||||||  |||||| || | ||||
Db      57580 TTTTTTTTGTATTTTTAATAGAGATGGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAAC 57521

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              |||| | ||   | |||  |||| ||  |||||||||||| ||||||||||||  | ||
Db      57520 TCCTGACCTCGTCGATCCGCCTGCCTTGGCCTCCCAAAATGCTGGGATTACAGGCATGAG 57461

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTA 1287
              ||||  ||||||||||   |  |  |   ||||
```

```
Db      57460 CCACCACACCCGGCTGTATGAAGCATGTTTTTA 57428


RESULT 14
US-09-984-890-3/c
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001306
;  CURRENT APPLICATION NUMBER: US/09/984,890
;  CURRENT FILING DATE:  2001-10-31
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 75395
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(75395)
;   OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

  Query Match             4.1%;  Score 85.8;  DB 4;  Length 75395;
  Best Local Similarity   78.6%;  Pred. No. 1.5e-11;
  Matches  114;  Conservative    1;  Mismatches   29;  Indels    1;  Gaps    1;

Qy       1125 TTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCT 1184
              | :|  | ||| ||| |  | ||| | |||||||||| || ||||||| ||||||||||
Db      44960 TAGTTTTTTATTTATTTCATTTTTTTTGTAGAGATGGCATCTTGCCATCTTGCCCAGGCT 44901

Qy       1185 AGTATTGAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATT 1243
              |  | |||||||| ||| |||||||||||||| ||| |||||||||| || |||||||
Db      44900 GGCCTCAAACTCCTAGGCTCAAGCAATCTTCCTGCCTCAGCCTCCCAAAGTGCTGGGATT 44841

Qy       1244 ACAGGTGTAAGCCACTGCACCCGGC 1268
              ||||| || ||||||  ||||| ||
Db      44840 ACAGGCGTGAGCCACCACACCCAGC 44816


RESULT 15
US-09-759-359A-3/c
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:
;  APPLICANT: ABU-THREIDEH, Jane et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001043
;  CURRENT APPLICATION NUMBER: US/09/759,359A
;  CURRENT FILING DATE:  2001-01-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 90541
;   TYPE: DNA
;   ORGANISM: Human
US-09-759-359A-3

  Query Match             4.1%;  Score 85.2;  DB 4;  Length 90541;
  Best Local Similarity   76.0%;  Pred. No. 2.2e-11;
  Matches  117;  Conservative    1;  Mismatches   35;  Indels    1;  Gaps    1;

Qy       1116 GTCTCATTTTTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTT 1175
              | || ||||||:|  | | ||| |||    |||   |||||||||||||| ||||||||
Db      23174 GGCTAATTTTTGTGTTTTTTATTTATTTATTTTAAGTAGAGATGGGATTTGGCCATGTT 23115

Qy       1176 GCCCAGGCTAGTATTGAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAAT 1234
              | ||||||| || |  ||||||||  || ||||||||  |||| ||  ||||||||||
Db      23114 GGCCAGGCTGGTCTCAAACTCCTAGCCTCAAGCAATCCACCTGCCTTGGCCTCCCAAAGT 23055

Qy       1235 GTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 1268
              | |||||||||||| || |||||| |  || |||
Db      23054 GCTGGGATTACAGGCGTGAGCCACAGTGCCTGGC 23021


Search completed: November 18, 2003, 07:06:35
Job time : 139.734 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:33:10 ; Search time 992.544 Seconds
(without alignments)
6829.176 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagttacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2071.6	99.9	2074	14	US-10-005-337A-2	Sequence 2, Appli
	2	367.8	17.7	2358	14	US-10-005-337A-1	Sequence 1, Appli
	3	144.4	7.0	1889	10	US-09-974-298-182	Sequence 182, App
	4	144.4	7.0	1889	12	US-10-240-965-158	Sequence 158, App
	5	144.4	7.0	1889	12	US-10-252-157-421	Sequence 421, App
	6	141.4	6.8	1325	14	US-10-084-817-217	Sequence 217, App
	7	129.4	6.2	184	9	US-09-923-876-5218	Sequence 5218, Ap
c	8	95.4	4.6	6040	9	US-09-764-860-1029	Sequence 1029, Ap
c	9	95.4	4.6	6040	14	US-10-074-095-1029	Sequence 1029, Ap
	10	94.8	4.6	300000	14	US-10-262-552-33	Sequence 33, Appl
c	11	93.8	4.5	119040	11	US-09-911-077A-19	Sequence 19, Appl
	12	93.2	4.5	32220	10	US-09-764-877-3933	Sequence 3933, Ap
	13	92.6	4.5	555	12	US-10-027-632-95562	Sequence 95562, A
	14	92.6	4.5	555	12	US-10-027-632-95563	Sequence 95563, A
	15	92.6	4.5	555	13	US-10-027-632-95562	Sequence 95562, A
	16	92.6	4.5	555	13	US-10-027-632-95563	Sequence 95563, A
	17	92.2	4.4	66804	9	US-09-740-041-3	Sequence 3, Appli
	18	92.2	4.4	66804	12	US-10-389-967-3	Sequence 3, Appli
c	19	91	4.4	461	11	US-09-918-995-35181	Sequence 35181, A
c	20	91	4.4	763	12	US-10-027-632-27383	Sequence 27383, A
c	21	91	4.4	763	12	US-10-027-632-27384	Sequence 27384, A
c	22	91	4.4	763	13	US-10-027-632-27383	Sequence 27383, A
c	23	91	4.4	763	13	US-10-027-632-27384	Sequence 27384, A
	24	91	4.4	56737	10	US-09-782-378A-17	Sequence 17, Appl
c	25	90.4	4.4	19820	10	US-09-764-877-2713	Sequence 2713, Ap
	26	90.2	4.3	701	12	US-10-027-632-113259	Sequence 113259,
	27	90.2	4.3	701	12	US-10-027-632-113260	Sequence 113260,
	28	90.2	4.3	701	13	US-10-027-632-113259	Sequence 113259,
	29	90.2	4.3	701	13	US-10-027-632-113260	Sequence 113260,
c	30	90.2	4.3	705	12	US-10-027-632-268052	Sequence 268052,
c	31	90.2	4.3	705	12	US-10-027-632-268053	Sequence 268053,
c	32	90.2	4.3	705	13	US-10-027-632-268052	Sequence 268052,
c	33	90.2	4.3	705	13	US-10-027-632-268053	Sequence 268053,
c	34	90	4.3	625	12	US-10-027-632-102687	Sequence 102687,
c	35	90	4.3	625	12	US-10-027-632-102688	Sequence 102688,
c	36	90	4.3	625	12	US-10-027-632-102689	Sequence 102689,
c	37	90	4.3	625	13	US-10-027-632-102687	Sequence 102687,
c	38	90	4.3	625	13	US-10-027-632-102688	Sequence 102688,
c	39	90	4.3	625	13	US-10-027-632-102689	Sequence 102689,
c	40	90	4.3	715	12	US-10-027-632-110575	Sequence 110575,
c	41	90	4.3	715	12	US-10-027-632-110576	Sequence 110576,
c	42	90	4.3	715	13	US-10-027-632-110575	Sequence 110575,
c	43	90	4.3	715	13	US-10-027-632-110576	Sequence 110576,
	44	89.8	4.3	86592	12	US-10-211-160-1	Sequence 1, Appli
	45	89.8	4.3	86592	12	US-10-051-681A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-005-337A-2
; Sequence 2, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; TITLE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-005-337A-2

Query Match 99.9%; Score 2071.6; DB 14; Length 2074;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCAGCAAGTTACTTAATGTTTTTTGCCTCAGCATCCTCTCTGTAAAATGAGAGCATTA 60

Qy         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTCATATCCAAGACCCTTAAA 120

Qy        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CATCCCACAGTCCTTCCCCCAAACACTTCTCCTCCTAATACCTCCCTCAGTTTGGGTCAG 180
```

```
Qy       181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240
Db       181 GCCTGGAACAAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240

Qy       241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300
Db       241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAA 300

Qy       301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
Db       301 ATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGT 360

Qy       361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420
Db       361 GAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAG 420

Qy       421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480
Db       421 AAAATAGTGATAGCTGTGTCCATCTCAAAGAAAAGCCCAGGAGATTTCCTTTATTTACCC 480

Qy       481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540
Db       481 CCTTTAAGATAGAATATTAGGAGACCGGAACATATGATACAGGAGGTACTGGGAGGGTCC 540

Qy       541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600
Db       541 CTCTTTGTCAATGTTTTGTCTTGGGGTGGGGAGTCGATGTCTTCTCAAAGTTTCAGAAAC 600

Qy       601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660
Db       601 ACCATCCACTGACTGAGCATTCAAGGGGCAAGAGGAGAATGGCAGCCACATTTGTTGATT 660

Qy       661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720
Db       661 GGGTGAGTTTGGGGAGAAATAGACACACAAAGGTCAAACATAACTTCCTAATTAACACTT 720

Qy       721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780
Db       721 CCCTCCATTCACAATTCCCTTCTCCCATTCTTCTCTCCTGTCTTTTACTSAKARAAACCC 780

Qy       781 AGTTTTTCCTGAAACTATAAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840
Db       781 AGTTTTTCCTGAAACTATAAAAAATACCCCCAGTATGTTTACATAATTTACACCTCAAAGA 840

Qy       841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900
Db       841 TTAGAAACCAGAAATAGAGACCTTTTCAACCCTTCCGGAAGCAAAGTGCATTATCCCTCC 900

Qy       901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960
Db       901 AGCCACGTGTCTCAAATCTTGATGCATCAGAATCATCTGGGTGCTTTKAAATTCAAGATG 960

Qy       961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020
Db       961 ATTCCTACGAGTTACCATAAATCAACTCAGAATTCCCTGGAGTGGGGCCAGGGATCTGTA 1020

Qy      1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
Db      1021 TTTCTGACAAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080

Qy      1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140
Db      1081 TGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATA 1140

Qy      1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200
Db      1141 TAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAA 1200

Qy      1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260
Db      1201 GCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTG 1260

Qy      1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
Db      1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320

Qy      1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380
Db      1321 GTAAAAATGCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATG 1380

Qy      1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440
Db      1381 GCCAGTGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATG 1440

Qy      1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500
Db      1441 GTAAGGTGGCATGTCATGGGGCCTATTTAGCCCAGACATCACTCCAAAGAATTCCAAACA 1500

Qy      1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560
Db      1501 GATATAGACAAGTGCCTTTAGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAA 1560

Qy      1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620
Db      1561 TAGGATGTCCTGGGACAAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAA 1620

Qy      1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
Db      1621 GATATTACTGGGGGTGTGATATGTAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680

Qy      1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740
Db      1681 AAGCTGACAAAGAAAAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCT 1740

Qy      1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800
Db      1741 GACTCTTGACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATG 1800

Qy      1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860
Db      1801 GCCCTCTCACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCC 1860

Qy      1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920
Db      1861 TCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCC 1920

Qy      1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980
Db      1921 TAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCA 1980

Qy      1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040
Db      1981 GCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCA 2040

Qy      2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
Db      2041 CGACAGAAAAACATACAAGACTCCTTCAGCCAAC 2074
```

RESULT 2
US-10-005-337A-1
; Sequence 1, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; TITLE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07

```
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2358
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-005-337A-1

  Query Match             17.7%;  Score 367.8;  DB 14;  Length 2358;
  Best Local Similarity   76.4%;  Pred. No. 1.2e-83;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348
              ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597

Qy       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405
              |||||||||||| ||| |||  ||  |||   |||||||||||||.|||| |   |||||
Db       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657

Qy       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465
              |||||||||||||||||||| ||  ||||| ||||| | |||||||||  |||||
Db       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717

Qy       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524
              ||||||  |||||||||||||| |||||||||||||||| |||||||| |     | |
Db       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777

Qy       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTT 1581
              |||  ||| | |||     |    ||  |||  | ||||||||||||     |  |||  |
Db       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837

Qy       1582 TCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
                |  || || |||| |||||||||   |||||||||||||||  |||   |||||||
Db       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897

Qy       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698
                || ||   | ||||||||||||| | ||||||||||| ||||||||||||  ||||
Db       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955

Qy       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757
              || |||| |||||||||||||| | ||||| |||||||||||||| | ||  || |  | ||
Db       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015

Qy       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817
              |||||| |||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075

Qy       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877
              |||||||| ||||  | || |      |   |  ||  |||||||||  ||||||||||
Db       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130

Qy       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937
              ||| |||||||||||||||||||||||||||||| |||||   ||  ||||||| || |||
Db       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190

Qy       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997
              | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249

Qy       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056
              |||||||||||| |  ||||||||  ||||||| ||| | | || || ||||||||||
Db       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309

Qy       2057 A 2057
              |
Db       2310 A 2310


RESULT 3
US-09-974-298-182
; Sequence 182, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, Huei-Mei
;  TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
;  FILE REFERENCE: PA-0037 P
;  CURRENT APPLICATION NUMBER: US/09/974,298
;  CURRENT FILING DATE:  2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,331
;  PRIOR FILING DATE: 2000-05-10
;  NUMBER OF SEQ ID NOS: 194
;  SOFTWARE: PERL Program
; SEQ ID NO 182
;   LENGTH: 1889
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020156263A1 332919.4
;   NAME/KEY: unsure
;   LOCATION: 1882
;   OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-182

  Query Match              7.0%;  Score 144.4;  DB 10;  Length 1889;
  Best Local Similarity   99.3%;  Pred. No. 2.6e-26;
  Matches  145;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1929 GCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 1988
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 64

Qy       1989 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 2048
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 124

Qy       2049 AAACATACAAGACTCCTTCAGCCAAC 2074
              |||||||||||||||||||||||||
Db        125 AAACATACAAGACTCCTTCAGCCAAC 150


RESULT 4
US-10-240-965-158
; Sequence 158, Application US/10240965
; Publication No. US20030165924A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  SHIFFMAN, Dov
;  APPLICANT:  SOMOGYI, Roland
;  APPLICANT:  LAWN, Richard M.
;  APPLICANT:  SEILHAMER, Jeffrey J.
;  APPLICANT:  PORTER, Gordon J.
;  APPLICANT:  MIKITA, Thomas
;  APPLICANT:  TAI, Julie
;  TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
;  FILE REFERENCE: PA-0025 PCT
;  CURRENT APPLICATION NUMBER: US/10/240,965
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: 60/195,106
;  PRIOR FILING DATE: 2000-04-05
;  NUMBER OF SEQ ID NOS: 276
;  SOFTWARE: PERL Program
; SEQ ID NO 158
;   LENGTH: 1889
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030165924A1 332919.4
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1655-1681, 1882
;   OTHER INFORMATION: a, t, c, g, or other
US-10-240-965-158
```

```
  Query Match             7.0%;  Score 144.4;  DB 12;  Length 1889;
  Best Local Similarity   99.3%;  Pred. No. 2.6e-26;
  Matches  145;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1929 GCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 1988
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 64

Qy       1989 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 124

Qy       2049 AAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||||
Db        125 AAACATACAAGACTCCTTCAGCCAAC 150


RESULT 5
US-10-252-157-421
; Sequence 421, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
;  APPLICANT: Faris, Mary
;  APPLICANT:  Pearson, Cecelia I.
;  TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
;  FILE REFERENCE: PA-0027-1 US
;  CURRENT APPLICATION NUMBER: US/10/252,157
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: 60/295,048
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 501
;  SOFTWARE: PERL Program
; SEQ ID NO 421
;   LENGTH: 1889
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030190640A1 332919.4
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1655-1681, 1882
;   OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-421

  Query Match             7.0%;  Score 144.4;  DB 12;  Length 1889;
  Best Local Similarity   99.3%;  Pred. No. 2.6e-26;
  Matches  145;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1929 GCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 1988
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATA 64

Qy       1989 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAA 124

Qy       2049 AAACATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||||
Db        125 AAACATACAAGACTCCTTCAGCCAAC 150


RESULT 6
US-10-084-817-217
; Sequence 217, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
;  APPLICANT: Susan Stuart
;  APPLICANT:  Jed G. Nuchtern
;  APPLICANT:  Sharon E. Plon
;  APPLICANT:  Jason M. Shohet
;  TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
;  FILE REFERENCE: PA-0046 US
;  CURRENT APPLICATION NUMBER: US/10/084,817
;  CURRENT FILING DATE:  2002-02-25
;  PRIOR APPLICATION NUMBER: 60/270,784
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 365
;  SOFTWARE: PERL Program
; SEQ ID NO 217
;   LENGTH: 1325
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030119009A1 183198CB1
;   NAME/KEY: unsure
;   LOCATION: 1258
;   OTHER INFORMATION: a, t, c, g, or other
US-10-084-817-217

  Query Match             6.8%;  Score 141.4;  DB 14;  Length 1325;
  Best Local Similarity   99.3%;  Pred. No. 1.3e-25;
  Matches  142;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1932 GGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGC 1991
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 GGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATGTCACCCGGCCCCAGCTATATAAGC 60

Qy       1992 TGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAA 2051
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAA 120

Qy       2052 CATACAAGACTCCTTCAGCCAAC 2074
             |||||||||||||||||||||||||
Db        121 CATACAAGACTCCTTCAGCCAAC 143


RESULT 7
US-09-923-876-5218
; Sequence 5218, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 5218
;   LENGTH: 184
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700456617H1
;   NAME/KEY: unsure
;   LOCATION: 67, 74, 98, 106, 122, 129
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-5218

  Query Match             6.2%;  Score 129.4;  DB 9;  Length 184;
  Best Local Similarity   95.3%;  Pred. No. 4.6e-23;
  Matches  141;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       1928 GGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATAT 1987
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 GGCGGGAGTGTTACTTCGGTTCCCAGGTTTGAAGATTATCTCACCCGGCCCCAGCTATAT 60
```

```
Qy       1988 AAGCTG-ACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAG 2046
              |||||| |||||| |||||||||||||||||||||| ||||||| |||||||||||||
Db         61 AAGCTGNACCGGTNTGGAGGGGCCCAGCAGGGCCAACNCCAGGGANTCCTTCCACGACAG 120

Qy       2047 AAAAACATACAAGACTCCTTCAGCCAAC 2074
              | |||||| ||||||||||||||||||||
Db        121 ANAAACATNCAAGACTCCTTCAGCCAAC 148


RESULT 8
US-09-764-860-1029/c
; Sequence 1029, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008
;  CURRENT APPLICATION NUMBER: US/09/764,860
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1198
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1029
;   LENGTH: 6040
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-860-1029

  Query Match             4.6%;  Score 95.4;  DB 9;  Length 6040;
  Best Local Similarity   72.8%;  Pred. No. 2.1e-13;
  Matches  123;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
              || | ||   ||| | || |||||||| ||| ||||| ||| ||||||| || |||||||
Db       3417 ATTTTTATATTTTTTAGTGGAGATGGGGTTTCGCCATATTGGCCAGGCTGGTCTTGAACT 3358

Qy       1196 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 1255
              ||| | || ||| |||| |||| ||| ||||| |||  || ||||||||||| || |||
Db       3357 CCTGACCTCAGCGATCTGCCTGCCTCAGCCTCTCAATGTGCTGGGATTACAGGCGTGAGC 3298

Qy       1256 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCAC 1304
              ||||||||| ||| | |   | ||   |||  | | ||| |||   ||
Db       3297 CACTGCACCTGGCCGTTTTTTTTTTCTTTTTCTTTTTTTTTTTGGAGAC 3249


RESULT 9
US-10-074-095-1029/c
; Sequence 1029, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008C1
;  CURRENT APPLICATION NUMBER: US/10/074,095
;  CURRENT FILING DATE:  2002-02-14
;  PRIOR APPLICATION NUMBER: 09/764,860
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/225,757
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/226,868
;  PRIOR FILING DATE: 2000-08-22
;  PRIOR APPLICATION NUMBER: 60/216,647
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/225,267
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/216,880
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/225,270
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/251,869
;  PRIOR FILING DATE: 2000-12-08
;  PRIOR APPLICATION NUMBER: 60/235,834
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: 60/234,274
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: 60/234,223
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: 60/228,924
;  PRIOR FILING DATE: 2000-08-30
;  PRIOR APPLICATION NUMBER: 60/224,518
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/236,369
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/224,519
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,964
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/241,809
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/249,299
;  PRIOR FILING DATE: 2000-11-17
;  PRIOR APPLICATION NUMBER: 60/236,327
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/241,785
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/244,617
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: 60/225,268
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/236,368
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/251,856
;  PRIOR FILING DATE: 2000-12-08
;  PRIOR APPLICATION NUMBER: 60/251,868
;  PRIOR FILING DATE: 2000-12-08
;  PRIOR APPLICATION NUMBER: 60/229,344
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/234,997
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: 60/229,343
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/229,345
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/229,287
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/229,513
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: 60/231,413
;  PRIOR FILING DATE: 2000-09-08
;  PRIOR APPLICATION NUMBER: 60/229,509
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: 60/236,367
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/237,039
;  PRIOR FILING DATE: 2000-10-02
;  PRIOR APPLICATION NUMBER: 60/237,038
;  PRIOR FILING DATE: 2000-10-02
;  PRIOR APPLICATION NUMBER: 60/236,370
```

; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match 4.6%; Score 95.4; DB 14; Length 6040;
Best Local Similarity 72.8%; Pred. No. 2.1e-13;
Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy       1136 ATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACT 1195
              || | ||   ||| | || |||||||| ||| ||||| ||| ||||||| || |||||||
Db       3417 ATTTTTATATTTTTTAGTGGAGATGGGGTTTCGCCATATTGGCCAGGCTGGTCTTGAACT 3358

Qy       1196 CCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGC 1255
              ||| | || ||| |||| |||| ||| ||||| |||  || ||||||||||||| || |||
Db       3357 CCTGACCTCAGCGATCTGCCTGCCTCAGCCTCTCAATGTGCTGGGATTACAGGCGTGAGC 3298

Qy       1256 CACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCAC 1304
              ||||||||| ||| | |   | ||   |||  | | ||| |||   ||
Db       3297 CACTGCACCTGGCCGTTTTTTTTTTCTTTTTCTTTTTTTTTTTGGAGAC 3249
```

RESULT 10
US-10-262-552-33
; Sequence 33, Application US/10262552
; Publication No. US20030125289A1
; GENERAL INFORMATION:
; APPLICANT: Gelb, Bruce D.
; APPLICANT: Tartaglia, Marco
; TITLE OF INVENTION: NOONAN SYNDROME GENE
; FILE REFERENCE: 2420/1J859-US1
; CURRENT APPLICATION NUMBER: US/10/262,552
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 60/326,532
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 300000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(300000)
; OTHER INFORMATION: where n may be a or g or c or t/u, unknown or other
US-10-262-552-33

Query Match 4.6%; Score 94.8; DB 14; Length 300000;
Best Local Similarity 77.1%; Pred. No. 3.3e-12;

```
  Matches  128;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | |||||| | |||||||||||| |||  |||||||| ||||||| || ||||||
Db     130900 TAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTTTTGAAC 130959

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
             |||| |  | | |  |||| |||| ||| |||||||||| || ||||||||||||||| |
Db     130960 TCCTGACTTCAGGTGATCTACCTGCCTCAGCCTCCCAAAGTGATGGGATTACAGGTGTGA 131019

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTG 1299
             ||||| ||| |||||  | |  | ||||  ||| |||| ||| |||
Db     131020 GCCACCGCAGCCGGCCAAAACTTTGTTTTTTTCCTCTTTTGTTG 131065


RESULT 11
US-09-911-077A-19/c
; Sequence 19, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
;  APPLICANT: BLAKELY, RANDY D.
;  APPLICANT:  APPARSUNDARAM, SUBRAMANIAM
;  APPLICANT:  FERGUSON, SHAWN
;  TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
;  FILE REFERENCE: VBLT:008US
;  CURRENT APPLICATION NUMBER: US/09/911,077A
;  CURRENT FILING DATE:  2001-07-23
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 119040
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (2347)..(90873)
;   OTHER INFORMATION: N = A, C, G or T/U
US-09-911-077A-19

  Query Match             4.5%;  Score 93.8;  DB 11;  Length 119040;
  Best Local Similarity   66.7%;  Pred. No. 3.4e-12;
  Matches  134;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       1140 ATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTA 1199
             || |||||| | |||||||||||||||| |||||||||||||| | |  ||||||||||
Db      85255 ATTGTATTTTTTGTAGAGATGGGATTTTTCCATGTTGCCCAGGGTGGCCTTGAACTCCTG 85196

Qy       1200 AGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACT 1259
                || ||||| |||||  || |||||||||| || ||||||| ||||||| |||||||
Db      85195 GCTCAAACAATCCTCCTGCTTCAGCCTCCCAAAGTGCTGGGATTGCAGGTGTGAGCCACT 85136

Qy       1260 GCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGA 1319
              | ||| ||  | |      |    || || || || | |      | | ||   |||
Db      85135 GTGCCCAGCCTACATACCTTGGTCTTGACCCTTTTCCATATTTTATTTTATTTTATTTTT 85076

Qy       1320 AGTAAAAATGCTCCAATTATT 1340
                | | | ||||    | ||
Db      85075 GAGACAGAGGCTCGCTCTGTT 85055


RESULT 12
US-09-764-877-3933
; Sequence 3933, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3933
;   LENGTH: 32220
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9867)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-3933

  Query Match             4.5%;  Score 93.2;  DB 10;  Length 32220;
  Best Local Similarity   79.2%;  Pred. No. 2.2e-12;
  Matches  122;  Conservative    1;  Mismatches   30;  Indels    1;  Gaps    1;

Qy       1117 TCTCATTTTTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTG 1176
             | |  |||||:|  | | | | | | |||||| | |||||||||||| |||||| ||||||
Db      30293 TTTTTTTTTTTTTTTTTTTTTTTTTTTGTATTTTTAGTAGAGATGGGGTTTTGCTATGTTG 30352

Qy       1177 CCCAGGCTAGTATTGAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATG 1235
             |||||||| || |||||||||| |||| |||  |||| |||| |||||| ||| ||
Db      30353 CCCAGGCTGGTCTTGAACTCCTTAGCTCAAGTGATCAGCCTGCCTCAGCCTCCTAAAGTG 30412

Qy       1236 TTGGGATTACAGGTGTAAGCCACTGCACCCGGCT 1269
              || |||||| |||||| |||||||||||| ||||
Db      30413 TTAGGATTATAGGTGTGAGCCACTGCACCTGGCT 30446


RESULT 13
US-10-027-632-95562
; Sequence 95562, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95562
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-95562

  Query Match             4.5%;  Score 92.6;  DB 12;  Length 555;
  Best Local Similarity   73.0%;  Pred. No. 2.6e-13;
  Matches  119;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
             ||  | | |||||| |||||||||  || |||  |||||||||||||||| || | ||||
Db        320 TAATTTTTGTATTTTTGGTAGAGACAGGGTTTCACCATGTTGCCCAGGCTGGTTTCGAAC 379

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
             |||| |||| |   |     ||| |||||||||||||| |||||||||||||||| || ||
Db        380 TCCTGAGCTCAAGCAATACACTGCCTCTGCCTCCCAAAGTGTTGGGATTACAGGCGTGAG 439

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCT 1297
```

```
                  ||||||| | |||  | |  |  | |  |||  ||| ||| |
Db        440 TCACTGCATCTGGCCAAAATTTCTTGTTTTTTTGTCTTTTTTT 482


RESULT 14
US-10-027-632-95563
; Sequence 95563, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95563
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-95563

  Query Match             4.5%;  Score 92.6;  DB 12;  Length 555;
  Best Local Similarity   73.0%;  Pred. No. 2.6e-13;
  Matches  119;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
                ||  | | |||||| |||||||||  || |||  ||||||||||||||||| || | ||||
Db        320 TAATTTTTGTATTTTTGGTAGAGACAGGGTTTCACCATGTTGCCCAGGCTGGTTTCGAAC 379

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
                |||| |||| |   |     ||| |||||||||||||| |||||||||||||||| || ||
Db        380 TCCTGAGCTCAAGCAATACACTGCCTCTGCCTCCCAAAGTGTTGGGATTACAGGCGTGAG 439

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCT 1297
                ||||||| | |||  | |  |  | |  |||  ||| ||| |
Db        440 TCACTGCATCTGGCCAAAATTTCTTGTTTTTTTGTCTTTTTTT 482


RESULT 15
US-10-027-632-95562
; Sequence 95562, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95562
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-95562

  Query Match             4.5%;  Score 92.6;  DB 13;  Length 555;
  Best Local Similarity   73.0%;  Pred. No. 2.6e-13;
  Matches  119;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
                ||  | | |||||| |||||||||  || |||  ||||||||||||||||| || | ||||
Db        320 TAATTTTTGTATTTTTGGTAGAGACAGGGTTTCACCATGTTGCCCAGGCTGGTTTCGAAC 379

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
                |||| |||| |   |     ||| |||||||||||||| |||||||||||||||| || ||
Db        380 TCCTGAGCTCAAGCAATACACTGCCTCTGCCTCCCAAAGTGTTGGGATTACAGGCGTGAG 439

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCT 1297
                ||||||| | |||  | |  |  | |  |||  ||| ||| |
Db        440 TCACTGCATCTGGCCAAAATTTCTTGTTTTTTTGTCTTTTTTT 482


Search completed: November 18, 2003, 07:42:07
Job time : 998.544 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:10 ; Search time 4326.29 Seconds
(without alignments)
11651.427 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagttacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	461.2	22.2	522	28	AQ480395	AQ480395 RPCI-11-2
	2	260.6	12.6	586	28	AZ295534	AZ295534 RPCI-23-1
	3	241	11.6	735	9	AU139209	AU139209 AU139209
	4	181	8.7	371	13	BY091041	BY091041 BY091041
	5	158.4	7.6	353	13	BY079997	BY079997 BY079997
c	6	98.6	4.8	343	28	AQ584530	AQ584530 RPCI-11-4
c	7	98.6	4.8	827	14	CD244827	CD244827 AGENCOURT
	8	95.8	4.6	502	28	AQ207760	AQ207760 HS_3026_B
c	9	95	4.6	412	9	AA494075	AA494075 ng61c08.s
	10	94.2	4.5	464	9	AA917468	AA917468 ol51f11.s
	11	94.2	4.5	2772	11	BC038630	BC038630 Homo sapi
c	12	93.8	4.5	315	9	AI039619	AI039619 ox28f05.x
	13	93.6	4.5	369	10	BF738217	BF738217 CM3-KT003
c	14	93.6	4.5	464	10	BF829139	BF829139 MR1-HN007
	15	93.6	4.5	723	28	AZ335882	AZ335882 1M0065G12
c	16	93.2	4.5	488	10	BF415552	BF415552 UI-R-CA1-
c	17	93.2	4.5	532	29	BZ260574	BZ260574 CH230-285
c	18	92.8	4.5	477	28	AZ406305	AZ406305 1M0175C06
c	19	92.6	4.5	380	9	AA947380	AA947380 od86h04.s
	20	92.4	4.5	534	10	BF724783	BF724783 bx08f06.y
c	21	92.2	4.4	387	10	BF400140	BF400140 UI-R-CA1-
c	22	92.2	4.4	453	28	AQ187593	AQ187593 HS_3132_A
	23	92	4.4	563	12	BI438856	BI438856 ic27e07.x
c	24	92	4.4	567	12	BI439182	BI439182 ic27e07.y
	25	92	4.4	950	10	BG390793	BG390793 602416521
	26	91.8	4.4	451	28	AQ108862	AQ108862 CIT-HSP-2
c	27	91.6	4.4	419	9	AA468422	AA468422 nc78f02.r
	28	91.6	4.4	717	28	AQ343576	AQ343576 RPCI11-12
c	29	91.4	4.4	442	28	AQ219900	AQ219900 HS_3252_A
	30	91.4	4.4	662	28	AZ365914	AZ365914 1M0112J19
	31	91.4	4.4	2073	11	AF289601	AF289601 Homo sapi
	32	91.2	4.4	470	9	AI814961	AI814961 wk69b11.x
	33	91	4.4	357	9	AW002875	AW002875 wr02a09.x
	34	91	4.4	464	9	AA845825	AA845825 ak85f12.s
c	35	91	4.4	556	29	BZ606409	BZ606409 WHAAB93TR
	36	91	4.4	587	9	AV720761	AV720761 AV720761
	37	91	4.4	587	9	AV720765	AV720765 AV720765
	38	91	4.4	589	9	AV720514	AV720514 AV720514
c	39	91	4.4	639	28	AQ109988	AQ109988 CIT-HSP-2
c	40	91	4.4	988	10	BF791763	BF791763 602251956
	41	91	4.4	1017	12	BM472183	BM472183 AGENCOURT
c	42	90.8	4.4	284	14	T39841	T39841 ya11c09.r3
c	43	90.8	4.4	541	28	AQ895064	AQ895064 HS_3175_A
	44	90.8	4.4	821	14	CD520493	CD520493 AGENCOURT
	45	90.6	4.4	430	9	AI288531	AI288531 ql91a04.x

ALIGNMENTS

RESULT 1
AQ480395/c
LOCUS AQ480395 522 bp DNA linear GSS 23-APR-1999
DEFINITION RPCI-11-236B22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-236B22, genomic survey sequence.
ACCESSION AQ480395
VERSION AQ480395.1 GI:4662514
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 522)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-11-236B22.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```
          library availability, please contact Pieter de Jong
          (pieter@dejong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genet cs (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
          Seq primer: T7
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7590285"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-236B22"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      168 a    100 c    113 g    141 t
ORIGIN

  Query Match            22.2%;  Score 461.2;  DB 28;  Length 522;
  Best Local Similarity   97.7%;  Pred. No. 2.5e-19;
  Matches  510;  Conservative    1;  Mismatches    5;  Indels    6;  Gaps    4;

Qy        971 GTTACCATAAATCAACTC-AGAATTCCCTGGAGTGGGGC-CAGGGATCTGTATTTCTGAC 1028
              |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Db        522 GTTACCATAAATCAACTCAAGAATTCCCTGGAGTGGGGCACAGGGATCTGTATTTCTGAC 463

Qy       1029 AAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGAACTTCAGCTCAATGACCTAA 1088
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        462 AAGCTCCCACAGGTGATTCCTTTCCCCACAGCATTTGAGTACTTCAGCTCAATGACCTAA 403

Qy       1089 TCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATATAGTATTT 1148
              |||||||||||||||||||| ||||||||||||||:||||||||||||||||||||||
Db        402 TCAGAGTCCTGCCATTGCTAATAACTGGTCTCATTTTTTTCATATATATATATAGTATTT 343

Qy       1149 GTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAAGCTAAGCA 1208
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        342 TTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCATAAGCTAAGCA 283

Qy       1209 ATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 1268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 ATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTGCACCCGGC 223

Qy       1269 TGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAAT 1328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 TGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAAT 163

Qy       1329 GCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAG 1385
              |||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        162 GCTCCAATTATTATGCTGTTTTAGAACACGGTAAGCATGTCATGTGCTAATTATGGCCAG 103

Qy       1386 TGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAG 1445
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        102 TGACATCATAAAAGAAAAGTGCATTACTGAATGCTTTCAATTTCTTATAATGATGGTAAG 43

Qy       1446 GTGGCATGTCATGGGGCCTATTTAG-CCCAGACATCACTCCA 1486
              |||||||||||||||||||||||| ||||||||||||||||
Db         42 GTGGCATGTCATGGGGCCTATTTAGCCCCAGACATCACTCCA 1


RESULT 2
AZ295534
LOCUS       AZ295534                586 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-105P3.TJ RPCI-23 Mus musculus genomic clone RPCI-23-105P3,
            genomic survey sequence.
ACCESSION   AZ295534
VERSION     AZ295534.1  GI:9537319
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-105P3.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 105  row: P  column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-105P3"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase.  Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      141 a    148 c    154 g    143 t
ORIGIN

  Query Match            12.6%;  Score 260.6;  DB 28;  Length 586;
  Best Local Similarity   75.9%;  Pred. No. 1.8e-07;
  Matches  403;  Conservative    0;  Mismatches  114;  Indels   14;  Gaps    6;

Qy       1533 CTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTTTCCCCTAA 1589
              ||||||||| ||||| ||| ||   ||||||||||||     |  |||  |  |  | ||
Db         21 CTTCCCCTCGGGCTGATTATCCCCAGAATAGGATGTCCCAAAGCAACACTTCCCAGCCAA 80

Qy       1590 GTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATATGTAGGGC 1649
              ||  ||||  |||||||||    ||||||||||||||  |||   |||||||     || ||
Db         81 CTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGC 140

Qy       1650 ATCTACATTTTCTTGATAGGTAGTCATATGAAAGCTGACAAAGAA--AAAAAGGGCAGTG 1707
                | |||||||||||| ||      |   | ||||||||||  ||||||  |||| |
Db        141 --TTGCATTTTCTTGATACGTTAGTCATTATTATCTGACAAAGAAGGAAAAAGAGCAGCG 198

Qy       1708 ATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGGATGACTTGC 1766
              |||||||||||| | ||||| ||||||||||||| | ||  || |  | ||||||||| ||
Db        199 ATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGC 258

Qy       1767 ATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTTCCTGATTCA 1826
              |||||||||| |||| ||||  |||||||||| ||||||||||||||||||||||||||
Db        259 ATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCG 318

Qy       1827 CATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCT 1886
              ||||  | || |    |  |  ||  |||||||||  |||||||||||||| |||||
Db        319 CATACGCCGCGGC-----CAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCT 373
```

```
Qy       1887 GGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGG 1946
             ||||||||||||||||||||||||| |||||  ||  |||||||  ||  ||||  |||| |
Db        374 GGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGC 433

Qy       1947 TTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGG 2006
             ||||||||  |||||||||||||||||||  ||||  ||||||||||  | |  |  ||||||||||
Db        434 TTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGG 492

Qy       2007 GGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACA 2057
             |||  |  ||||||||  |||||||  |||  |  |          ||||||||  |
Db        493 GGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACATGAGAGAAAACATAGA 543
```

RESULT 3
AU139209

LOCUS	AU139209 735 bp mRNA linear EST 02-AUG-2002
DEFINITION	AU139209 PLACE1 Homo sapiens cDNA clone PLACE1010155 5', mRNA sequence.
ACCESSION	AU139209
VERSION	AU139209.1 GI:11000730
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 735)
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
TITLE	HRI human cDNA project
JOURNAL	Unpublished
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

```
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1010155"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
BASE COUNT      240 a    154 c    181 g    156 t      4 others
ORIGIN

  Query Match          11.6%;  Score 241;  DB 9;  Length 735;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-06;
  Matches  241;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1834 AGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCTGGAATGA 1893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 AGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCTGGAATGA 69

Qy       1894 AAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGGTTCCCAG 1953
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 AAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGGTTCCCAG 129

Qy       1954 GTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAG 2013
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAG 189

Qy       2014 CAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTTCAGCCAA 2073
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CAGGGCCAACTCCAGGGATTCCTTCCACGACAGAAAAACATACAAGACTCCTTCAGCCAA 249

Qy       2074 C 2074
             |
Db        250 C 250
```

RESULT 4
BY091041

LOCUS	BY091041 371 bp mRNA linear EST 07-DEC-2002
DEFINITION	BY091041 RIKEN full-length enriched, 10 days neonate heart Mus musculus cDNA clone K630084E21 5', mRNA sequence.
ACCESSION	BY091041
VERSION	BY091041.1 GI:26204320
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 371)
AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	22354683
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630084E21"
                     /tissue_type="heart"
                     /dev_stage="10 days neonate"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     heart"
BASE COUNT       91 a     94 c    105 g     81 t
ORIGIN

  Query Match             8.7%;  Score 181;  DB 13;  Length 371;
  Best Local Similarity   81.3%;  Pred. No. 0.012;
  Matches  248;  Conservative    0;  Mismatches   50;  Indels    7;  Gaps    3;

Qy       1754 TAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATT 1813
             | |||||||| |||||||||||||| |||| ||||  |||||||||||| ||||||||||||
Db          5 TGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATT 64

Qy       1814 TCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCC 1873
             ||||||||||||  ||||   | ||  |       |    |  ||  |||||||||  |||||||
Db         65 TCTTCCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCC 119

Qy       1874 AGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGG 1933
             |||||||  |||||||||||||||||||||||||||||  |||||   ||  |||||||  ||
Db        120 AGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGG 179

Qy       1934 AGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTG 1993
              ||||  |||| | ||||||| ||||||||||||||||||| ||||  |||||||||| | |
Db        180 GGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGG 238

Qy       1994 ACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAAC 2052
              | |||||||||||||| |  ||||||||  ||||||| ||| | | || || ||||||||
Db        239 GCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAAC 298

Qy       2053 ATACA 2057
              ||| |
Db        299 ATAGA 303


RESULT 5
BY079997
LOCUS       BY079997                353 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY079997 RIKEN full-length enriched, 10 days neonate heart Mus
            musculus cDNA clone K630022E19 5', mRNA sequence.
ACCESSION   BY079997
VERSION     BY079997.1  GI:26185447
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630022E19"
                     /tissue_type="heart"
                     /dev_stage="10 days neonate"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     heart"
BASE COUNT       82 a     88 c    102 g     79 t      2 others
ORIGIN

  Query Match             7.6%;  Score 158.4;  DB 13;  Length 353;
  Best Local Similarity   73.9%;  Pred. No. 0.27;
```

```
  Matches  229;  Conservative    0;  Mismatches   73;  Indels    8;  Gaps    2;

Qy       1748 GACAAATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCT 1807
              || |  | |||||||| |||||||||||| |||| ||||  |||||||||| ||||||
Db          2 GATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCT 61

Qy       1808 CACATTTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAG 1867
              |||||||||||||||||  |||  | || |      |   |  ||  |||||||||  |
Db         62 CACATTTCTTCCTGATTCGNATACGCCGCGG-----NCAGCTTGTCATCTCCCTCTTGGG 116

Qy       1868 CTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGG 1927
              |||||||||||||| |||||||||||||||||||||||||||||| |||||  ||  ||||
Db         117 CTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGG 176

Qy       1928 GGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATAT 1987
              ||| || |||| |||| | |||||||| ||||||   || || | |   |||  |
Db         177 GGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAG---ATATCTCACCAGCCCTACCTATA 233

Qy       1988 AAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTCCACGACAGA 2047
              | | | | |||||||||||||| |  |||||||  ||  ||| |      || || |||
Db         234 TAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCAGGGGTTCATCCACAAGAGAGA 293

Qy       2048 AAAACATACA 2057
              |||||||| |
Db         294 AAAACATAGA 303


RESULT 6
AQ584530/c
LOCUS       AQ584530                343 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  RPCI-11-456H3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-456H3,
            genomic survey sequence.
ACCESSION   AQ584530
VERSION     AQ584530.1  GI:5011640
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7674890"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-456H3"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      123 a     60 c     68 g     92 t
ORIGIN

  Query Match             4.8%;  Score 98.6;  DB 28;  Length 343;
  Best Local Similarity   72.1%;  Pred. No. 9.6e+02;
  Matches  142;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy       1096 CCTGCCATTGCTAATATCTGGTCTCATTTTTBTCATATATATATATAGTATTTGTGGTAG 1155
              | ||||       || ||| | | || ||||| | || | | |||   |||||| ||||||
Db        324 CATGCCCAGCTAAAGATCAGATATCCTTTTTATAATTTTTTTATTGTGTATTTTTGGTAG 265

Qy       1156 AGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAACTCCTAAGCT-AAGCAATCTTC 1214
              |||  || |||  |||||||| ||||||| || |||||||||| | || ||   |||| |
Db        264 AGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAAATGATCTGC 205

Qy       1215 CTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTGCACCCGGCTGATAG 1274
              |||  || | |||||||| || ||||||||||||||| ||||||||||| |||  |||
Db        204 CTGCATCGGTCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGCACCTGGCCCATAT 145

Qy       1275 CTGGTTTCATTTACTCT 1291
              ||  ||  |||  | |
Db        144 GTGTGTTGTTTTTTTTT 128


RESULT 7
CD244827/c
LOCUS       CD244827                827 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  AGENCOURT_14096429 NIH_MGC_181 Homo sapiens cDNA clone
            IMAGE:30376302 5', mRNA sequence.
ACCESSION   CD244827
VERSION     CD244827.1  GI:31005291
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM438  row: 1  column: 07
            High quality sequence stop: 638.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30376302"
                     /tissue_type="White Matter"
                     /dev_stage="Unknown"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_181"
                     /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV
                     (destroyed); Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.42 kb. Library was constructed by
                     (Invitrogen). Note: this is a NIH_MGC Library."
BASE COUNT      219 a    159 c    156 g    291 t      2 others
ORIGIN

  Query Match             4.8%;  Score 98.6;  DB 14;  Length 827;
  Best Local Similarity   74.0%;  Pred. No. 5.5e+02;
  Matches  125;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        258 AAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAAATAGGAGCTATACAAAG 317
```

```
              ||  | |||  ||  || | ||||| |||| ||||||||||||||| ||| | ||||| ||
Db        582 AATAACTAAGCACGGTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAG 523

Qy        318 AAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGTGAAACATTCCATATATT 377
              ||||||||||||| | ||| ||||| ||||||| |||||| ||||| | ||||||||  |
Db        522 AAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTA 463

Qy        378 AAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAGAAAATA 426
              ||||| ||| ||| |  ||  | ||    |||||   || || | || |
Db        462 AAAAAAAAAAAAAAAGAAACAAAAACCACAAAAACAAAACAAAACAAGA 414


RESULT 8
AQ207760
LOCUS       AQ207760               502 bp    DNA     linear   GSS 18-SEP-1998
DEFINITION  HS_3026_B1_G10_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3026 Col=19 Row=N, genomic survey
            sequence.
ACCESSION   AQ207760
VERSION     AQ207760.1  GI:3620495
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3026  row: N  column: 19
            Class: BAC ends
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3026 Col=19 Row=N"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      130 a    117 c    110 g    140 t      5 others
ORIGIN

  Query Match             4.6%;  Score 95.8;  DB 28;  Length 502;
  Best Local Similarity   71.9%;  Pred. No. 1.1e+03;
  Matches  138;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | | |||||| | ||||| ||||| ||||||||||||||||||||| || ||||||
Db        291 TAATTTTTGTATTTTTTGTAGACATGGGTTTTTGCCATGTTGCCCAGGCTGGTCTTGAAC 350

Qy       1195 TCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
              |||| |||| ||||| ||| |||| ||| ||||| |||| || ||||||||||||   |||
Db        351 TCCTGAGCTCAAGCAGTCTGCCTGCCTCAGCCTCGCAAAGTGCTGGGATTACAGGCATAA 410

Qy       1254 GCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCA 1313
              |||||||||||| ||    ||  |  |     ||| | ||||| |  |      | ||
Db        411 GCCACTGCACCCAGCCAGGAGAAAGAATNTAAAACTATCTTTCTGTAAGAAAAAAGTACA 470

Qy       1314 TTTTGAAGTAAA 1325
              ||    |  ||||
Db        471 TTGCAACATAAA 482


RESULT 9
AA494075/c
LOCUS       AA494075               412 bp    mRNA    linear   EST 19-AUG-1997
DEFINITION  ng61c08.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:939278
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA494075
VERSION     AA494075.1  GI:2223916
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 841   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:939278"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lip2"
                     /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                     made by oligo-dT priming. Non- directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT      118 a     82 c    114 g     98 t
ORIGIN

  Query Match             4.6%;  Score 95;  DB 9;  Length 412;
  Best Local Similarity   74.8%;  Pred. No. 1.4e+03;
  Matches  119;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1129 CATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTA 1188
              | ||  |   | | ||||||| | |||||||||| |||||||||||| |||||| |  |
Db        272 CCTAGCTTATTTTTGTATTTTTTGTAGAGATGGGGTTTTGCCATGTTACCCAGGTTGATC 213

Qy       1189 TTGAACTCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGG 1248
              |  ||||||| |||| |||| ||| |||| ||| |||||||||| || ||||||||||||
Db        212 TCAAACTCCTGAGCTCAGCAGTCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGG 153

Qy       1249 TGTAAGCCACTGCACCCGGCTGATAGCTGGTTTCATTTA 1287
              || ||||||  |||| |||  ||| || |     ||||
Db        152 CGTGAGCCACCACACCTGGCCAATAACTTGCAAATTTTA 114


RESULT 10
AA917468
LOCUS       AA917468               464 bp    mRNA    linear   EST 10-JUN-1998
```

```
DEFINITION  ol51f11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1527021 3' similar to SW:TSG6_HUMAN P98066 TUMOR NECROSIS
            FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR ;contains Alu repetitive
            element;, mRNA sequence.
ACCESSION   AA917468
VERSION     AA917468.1  GI:3057358
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 417   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1527021"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      118 a    101 c     96 g    149 t
ORIGIN

  Query Match             4.5%;  Score 94.2;  DB 9;  Length 464;
  Best Local Similarity   70.4%;  Pred. No. 1.4e+03;
  Matches  126;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy       1117 TCTCATTTTTBTCATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTG 1176
              | | ||||||    | |  || || | | ||| |  |||||||| | | |||| ||||||
Db          1 TTTAATTTTTAATTTTTTAATTTAAAATTTTTTTCATAGAGATGAGGTCTTGCTATGTTG 60

Qy       1177 CCCAGGCTAGTATTGAACTCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGT 1236
              ||||| || || || ||||||| |||| |  |   |  ||| |||||||| ||||| ||
Db         61 CCCAGCCTTGTTTTTAACTCCTGAGCTCAAGAGATTCTCTGCCTCTGCCTTCCAAAGTGA 120

Qy       1237 TGGGATTACAGGTGTAAGCCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTT 1295
              ||||||||||||| || |||||||||  | | | |||   |||||||   | |||   |||
Db        121 TGGGATTACAGGAGTGAGCCACTGCGCGCCGGTGAATTCTGGTTTTTATGACTTGCTTT 179


RESULT 11
BC038630
LOCUS       BC038630                2772 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, Similar to hypothetical protein FLJ20489, clone
            IMAGE:5263792, mRNA.
ACCESSION   BC038630
VERSION     BC038630.1  GI:24116283
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2772)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: i Column: 4
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .2772
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5263792"
                     /tissue_type="Brain, hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
BASE COUNT      682 a    623 c    602 g    865 t
ORIGIN

  Query Match             4.5%;  Score 94.2;  DB 11;  Length 2772;
  Best Local Similarity   68.9%;  Pred. No. 4.8e+02;
  Matches  144;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    1;

Qy       1129 CATATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTA 1188
              ||  ||| | ||| | |||| | ||||||| ||| |||||||||||||| ||||||| ||
Db        997 CACCTATTTTTATTGCATTTTTAGTAGAGACGGGTTTTTGCCATGTTGGCCAGGCTGGTC 1056

Qy       1189 TTGAACTCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGG 1248
              || ||||||| | || ||    | ||| ||| |||||||||| || ||||||||||||
Db       1057 TTCAACTCCTGACCTCAGGTGAACTTCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGG 1116

Qy       1249 TGTAAGCCACTGCACCCGGCTGATA--GCTGGTTTCATTTACTCTATTTCTTGACCACTC 1306
              || ||||||  | ||||||  |||   | | ||| | | || | || |   | ||||
Db       1117 CGTGAGCCACCACGCCCGGCCTATAATTCAGTTTTAAATAACACCATCTGAAGGCCACCA 1176

Qy       1307 TGATCCATTTTGAAGTAAAAATGCTCCAA 1335
              ||  |   |   ||||  |||  |||||
Db       1177 AGAGTCCAGTCCCAGTAGGAATATTCCAA 1205


RESULT 12
AI039619/c
LOCUS       AI039619                315 bp    mRNA    linear   EST 30-JUN-1998
DEFINITION  ox28f05.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1657665 3' similar to WP:T20D3.3 CE03672 ;, mRNA sequence.
ACCESSION   AI039619
VERSION     AI039619.1  GI:3278813
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
```

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 296.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1657665"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       61 a     65 c     69 g    120 t
ORIGIN

  Query Match             4.5%;  Score 93.8;  DB 9;  Length 315;
  Best Local Similarity   72.2%;  Pred. No. 1.9e+03;
  Matches  122;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        258 AAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAAAATAGGAGCTATACAAAG 317
              ||  | |||  ||  || | ||||| |||| ||||||||||||||| ||| | ||||| ||
Db        177 AATAACTAAGCACGGTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAG 118

Qy        318 AAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTGTGAAACATTCCATATATT 377
              |||||||||||||| | ||| ||||| ||||||| |||||| ||||| | ||||||||
Db        117 AAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTA 58

Qy        378 AAAAATAAATAAATAATAAAGAGAAAAGGAAAAAATTAAAAAGAAAATA 426
              || || ||| ||| |  ||  | ||      |||   || || | || |
Db         57 AAGAAAAAAAAAAAAGAAACAAAAACCCCCCAAACAAAACAAAACAAGA 9


RESULT 13
BF738217
LOCUS       BF738217                369 bp    mRNA    linear   EST 10-JAN-2001
DEFINITION  CM3-KT0033-151200-572-f07 KT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF738217
VERSION     BF738217.1  GI:12064893
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-KT0033-
            151200-572-f07&t3=2000-12-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 369.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="KT0033"
                     /note="Organ: bladder_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       90 a     85 c     78 g    116 t
ORIGIN

  Query Match             4.5%;  Score 93.6;  DB 10;  Length 369;
  Best Local Similarity   67.3%;  Pred. No. 1.8e+03;
  Matches  132;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy       1135 TATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATTGAAC 1194
              ||  | | |||||| ||||| ||||||| |||  |||| ||| ||||||| || |||||
Db         79 TAATTTTTGTATTTTTGGTAAAGATGGGGTTTCACCATTTTGGCCAGGCTGGTCTTGAAT 138

Qy       1195 TCCTAAGCTAAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGTGTAAG 1254
              |||| | | |||  |||| |||| |||||||||||||| || ||||||||||||||| |
Db        139 TCCTGACCCAAGTGATCTCCCTGCCTCTGCCTCCCAAAGTGCTGGGATTACAGGTGTGAA 198

Qy       1255 CCACTGCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCCAT 1314
               |||| ||||| ||   ||| |   |  |  |||  | | |   |  ||| |
Db        199 TCACTACACCCAGCCTTTAGATTTGTGTAAGTACATTCTATGAGGTTCACACAATGATGA 258

Qy       1315 TTTGAAGTAAAAATGC 1330
                |    ||| |||||
Db        259 AATTGCTTAAGAATGC 274


RESULT 14
BF829139/c
LOCUS       BF829139                464 bp    mRNA    linear   EST 13-JAN-2001
DEFINITION  MR1-HN0070-151200-003-h04 HN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF829139
VERSION     BF829139.1  GI:12174344
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
```

```
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR1&t2=MR1-HN0070-
            151200-003-h04&t3=2000-12-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HN0070"
                     /note="Organ: head_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      138 a     95 c    145 g     86 t
ORIGIN

  Query Match             4.5%;  Score 93.6;  DB 10;  Length 464;
  Best Local Similarity   67.9%;  Pred. No. 1.6e+03;
  Matches  144;  Conservative    1;  Mismatches   66;  Indels    1;  Gaps    1;

Qy       1071 TTCAGCTCAATGACCTAATCAGAGTCCTGCCATTGCTAATATCTGGTCTCATTTTTBTCA 1130
              | || |    | | ||   || ||| |||  ||| |  |||  ||    |     : |
Db        406 TCCACCCGCCTTAGCTTCCCACAGTGCTGGGATTACAGATAGGTGTGAGCCACCGCGCCC 347

Qy       1131 TATATATATATAGTATTTGTGGTAGAGATGGGATTTTGCCATGTTGCCCAGGCTAGTATT 1190
                || || | |||||| | ||||||| ||| |||||||||||||||||||| || ||
Db        346 AGCCTAAATTTTGTATTTTTAGTAGAGACGGGGTTTTGCCATGTTGCCCAGGCTGGTCTT 287

Qy       1191 GAACTCCTAAGCT-AAGCAATCTTCCTGTCTCTGCCTCCCAAAATGTTGGGATTACAGGT 1249
              ||||||||| ||| |||  |||  ||   ||| |||||||||| ||||||||||||||||
Db        286 GAACTCCTAGGCTCAAGAGATCCACCCACCTCCGCCTCCCAAAGTGTTGGGATTACAGGT 227

Qy       1250 GTAAGCCACTGCACCCGGCTGATAGCTGGTTT 1281
              || |||||| ||||| |||   |  |  |||
Db        226 GTGAGCCACCGCACCTGGCCAGTTTATTCTTT 195


RESULT 15
AZ335882
LOCUS       AZ335882                723 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0065G12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0065G12 R, genomic survey sequence.
ACCESSION   AZ335882
VERSION     AZ335882.1  GI:10404635
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0065  row: G  column: 12
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0065G12"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      275 a    105 c    175 g    167 t      1 others
ORIGIN

  Query Match             4.5%;  Score 93.6;  DB 28;  Length 723;
  Best Local Similarity   71.1%;  Pred. No. 1.2e+03;
  Matches  123;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        240 AGATGAAGAGACCAATGAAAATAGTAATGACTCTGTTTGCTTCAGCAGGACATATACTAA 299
              |||  |||      ||| | ||  |  |  ||| || | ||||| |||| ||||||||||
Db        530 AGACAAAGCACATAATTANAAATGAGAGTACTGTGCTCGCTTCGGCAGCACATATACTAA 589

Qy        300 AATAGGAGCTATACAAAGAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 359
              ||| ||| | ||||| ||||||||||||||| | ||| ||||| ||||||| |||||| |
Db        590 AATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAATTCG 649

Qy        360 TGAAACATTCCATATATTAAAAATAAATAAATAATAAAGAGAAAAGGAAAAAA 412
              |||| | |||||||| ||  ||||| | ||   | || ||    |    ||||
Db        650 TGAAGCGTTCCATATTTTTGAAATACAAAATGCAAAAGGATCCTAACTCAAAA 702


Search completed: November 18, 2003, 07:01:59
Job time : 4336.79 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:30:20 ; Search time 8578.08 Seconds
(without alignments)
11245.517 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggatcctttcatgtttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2358	100.0	2358	6	AX468603	AX468603 Sequence
	2	2301.8	97.6	142902	2	AC119234	AC119234 Mus muscu
	3	879.8	37.3	229640	2	AC105469	AC105469 Rattus no
	4	879.8	37.3	238344	2	AC097115	AC097115 Rattus no
c	5	804.8	34.1	838	11	BV062893	BV062893 S212P6025
	6	646.2	27.4	723	10	AF478692	AF478692 Mus muscu
	7	640.2	27.2	665	11	BV076484	BV076484 S212P6036
	8	367.8	15.6	2074	6	AX468604	AX468604 Sequence
	9	367.8	15.6	2074	9	AF131884	AF131884 Homo sapi
c	10	347.6	14.7	50111	9	AL590622	AL590622 Human DNA
	11	128.6	5.5	1901	6	AX322775	AX322775 Sequence
	12	128.6	5.5	1901	6	BD094076	BD094076 Shear str
	13	128.6	5.5	1901	9	HSRNACINP	X83703 H.sapiens m
	14	128.6	5.5	1901	11	G28603	G28603 human STS S
c	15	127.2	5.4	160350	2	AC074094	AC074094 Homo sapi
	16	125.8	5.3	110480	10	AC122467	AC122467 Mus muscu
	17	125.8	5.3	169126	2	AC132348	AC132348 Mus muscu
c	18	125.2	5.3	152346	2	AC102022	AC102022 Mus muscu
c	19	125.2	5.3	258445	2	AC122205	AC122205 Mus muscu
c	20	122.4	5.2	158357	9	AL365434	AL365434 Human DNA
	21	121.6	5.2	1940	4	AF131883	AF131883 Oryctolag
c	22	119.6	5.1	25203	10	AL365324	AL365324 Mouse DNA
c	23	119.6	5.1	184754	2	AC022675	AC022675 Mus muscu
	24	119	5.0	234469	2	AC119697	AC119697 Rattus no
	25	117.8	5.0	153899	10	AL772285	AL772285 Mouse DNA
c	26	117.8	5.0	191485	2	AC091467	AC091467 Mus muscu
c	27	117.8	5.0	227165	10	AL626768	AL626768 Mouse DNA
c	28	117.8	5.0	243290	10	AL663088	AL663088 Mouse DNA
	29	117.8	5.0	317209	2	AL772175	AL772175 Mus muscu
c	30	116.6	4.9	167591	2	AC136093	AC136093 Rattus no
	31	116.6	4.9	179706	2	AC128045	AC128045 Rattus no
	32	116.6	4.9	231573	2	AC123610	AC123610 Mus muscu
c	33	116.6	4.9	245489	2	AC105665	AC105665 Rattus no
c	34	116.6	4.9	267971	2	AC128995	AC128995 Rattus no
c	35	116.4	4.9	183205	2	AC133734	AC133734 Rattus no
	36	116.4	4.9	238313	2	AC109696	AC109696 Rattus no
c	37	116.4	4.9	240006	2	AC094069	AC094069 Rattus no
	38	116.2	4.9	249982	2	AC108661	AC108661 Rattus no
c	39	115.8	4.9	167223	2	AC113514	AC113514 Mus muscu
	40	115.8	4.9	219619	2	AC116128	AC116128 Mus muscu
	41	115.4	4.9	244328	2	AC094938	AC094938 Rattus no
	42	115.2	4.9	171403	10	AC124423	AC124423 Mus muscu
c	43	114.8	4.9	294703	2	AC103396	AC103396 Mus muscu
c	44	114.6	4.9	176926	2	AC138341	AC138341 Mus muscu
c	45	114.4	4.9	167601	2	AC115886	AC115886 Mus muscu

ALIGNMENTS

RESULT 1
AX468603
LOCUS AX468603 2358 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0246220.
ACCESSION AX468603
VERSION AX468603.1 GI:21901402
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and uses thereof

```
  JOURNAL   Patent: WO 0246220-A 1 13-JUN-2002;
            Aventis Pharma S.A. (FR) ; The Regents of The University of
            California at San Diego (US) ; Benoit, Patrick (FR)
FEATURES             Location/Qualifiers
     source          1. .2358
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
BASE COUNT      636 a    567 c    542 g    613 t
ORIGIN

  Query Match             100.0%;  Score 2358;  DB 6;  Length 2358;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2358;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT  60

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180

Qy        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240

Qy        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300

Qy        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360

Qy        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420

Qy        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480

Qy        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540

Qy        541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600

Qy        601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660

Qy        661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720

Qy        721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780

Qy        781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840

Qy        841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900

Qy        901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960

Qy        961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020

Qy       1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080

Qy       1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140

Qy       1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200

Qy       1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260

Qy       1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320

Qy       1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380

Qy       1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440

Qy       1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500

Qy       1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560

Qy       1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620

Qy       1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680

Qy       1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740

Qy       1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800

Qy       1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860

Qy       1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920

Qy       1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
```

```
Qy       1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040

Qy       2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100

Qy       2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160

Qy       2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220

Qy       2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280

Qy       2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340

Qy       2341 GGTCGGAGGCCACCATGG 2358
             ||||||||||||||||||
Db       2341 GGTCGGAGGCCACCATGG 2358
```

```
RESULT 2.
AC119234
LOCUS       AC119234              142902 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION  Mus musculus clone RP24-211P24, *** SEQUENCING IN PROGRESS ***, 2
            ordered pieces.
ACCESSION   AC119234
VERSION     AC119234.9  GI:31581760
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 142902)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-211P24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142902)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 142902)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 11, 2003 this sequence version replaced gi:31455706.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25236
                Center clone name: 211_P_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    61126: contig of 61126 bp in length
            *   61127    61226: gap of 100 bp
            *   61227   142902: contig of 81676 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142902
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-211P24"
                     /clone_lib="RPCI-24 Male Mouse BAC"
BASE COUNT    41376 a  30899 c  29987 g  40473 t    167 others
ORIGIN
```

```
  Query Match             97.6%;  Score 2301.8;  DB 2;  Length 142902;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2314;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89306 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 89365

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89366 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 89425

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89426 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 89485
```

```
Qy       181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89486 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 89545

Qy       241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89546 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 89605

Qy       301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89606 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 89665

Qy       361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89666 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 89725

Qy       421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89726 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 89785

Qy       481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89786 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 89845

Qy       541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89846 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 89905

Qy       601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89906 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 89965

Qy       661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89966 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 90025

Qy       721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90026 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 90085

Qy       781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90086 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 90145

Qy       841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90146 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 90205

Qy       901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90206 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 90265

Qy       961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90266 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 90325

Qy      1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90326 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 90385

Qy      1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90386 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 90445

Qy      1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90446 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 90505

Qy      1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90506 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 90565

Qy      1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     90566 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTA-T 90624

Qy      1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90625 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 90684

Qy      1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90685 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 90744

Qy      1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90745 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 90804

Qy      1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90805 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 90864

Qy      1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90865 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 90924

Qy      1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90925 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 90984

Qy      1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90985 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 91044

Qy      1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91045 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 91104

Qy      1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91105 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 91164

Qy      1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91165 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 91224

Qy      1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91225 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 91284

Qy      1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91285 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 91344

Qy      2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91345 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 91404

Qy      2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91405 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 91464

Qy      2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91465 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 91524

Qy      2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91525 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 91584

Qy      2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGG 2317
            |||||||||||||||||||||||||||||||  ||
Db     91585 GGGTTCATCCACAAGAGAGAAAAACATAGACTCACGG 91621
```

```
RESULT 3
AC105469
LOCUS       AC105469               229640 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-140I18, WORKING DRAFT SEQUENCE, 9
            unordered pieces.
ACCESSION   AC105469
VERSION     AC105469.4  GI:25007338
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229640)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229640)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 229640)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23101653.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNAG
                Center clone name: CH230-140I18
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 215126 bases at least Q40
                Consensus quality: 217436 bases at least Q30
                Consensus quality: 218866 bases at least Q20
                Estimated insert size: 214300; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   215467: contig of 215467 bp in length
            *   215468   215567: gap of unknown length
            *   215568   218782: contig of 3215 bp in length
            *   218783   218882: gap of unknown length
            *   218883   220359: contig of 1477 bp in length
            *   220360   220459: gap of unknown length
            *   220460   221614: contig of 1155 bp in length
            *   221615   221714: gap of unknown length
            *   221715   223060: contig of 1346 bp in length
            *   223061   223160: gap of unknown length
            *   223161   224771: contig of 1611 bp in length
            *   224772   224871: gap of unknown length
            *   224872   225886: contig of 1015 bp in length
            *   225887   225986: gap of unknown length
            *   225987   228180: contig of 2194 bp in length
            *   228181   228280: gap of unknown length
            *   228281   229640: contig of 1360 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229640
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-140I18"
     misc_feature    1. .1268
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    4765. .7050
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(7251. .8156)
```

```
                    /note="clone_boundary
                    clone_end:T7
                    site:
                    end_sequence:RWAXF57TJB"
     misc_feature   complement(7252. .8118)
                    /note="clone_boundary
                    clone_end:T7
                    site:
                    end_sequence:RWAXF57TJC"
     misc_feature   213542. .215467
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   215568. .216791
                    /note="wgs_end_extension
                    clone_end:T7"
BASE COUNT    61330 a  47195 c  48097 g  63316 t   9702 others
ORIGIN

  Query Match            37.3%;  Score 879.8;  DB 2;  Length 229640;
  Best Local Similarity  82.3%;  Pred. No. 1e-245;
  Matches 1209;  Conservative    0;  Mismatches  177;  Indels   83;  Gaps   14;

Qy        879 ATGAAAACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
              ||| |||| ||||| ||||||||| ||| | ||||| |||||||||||  |||||||||
Db     140789 ATGGAAACCAGAAGGCCAACAGTTGTCCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 140848

Qy        939 AGAACACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGC 998
              |||||||||||| |||||||||||||     ||| |||||| |||||| ||||||
Db     140849 ----CACTAGGAGAGGAGAACCCACGAAGGA-----TATCAGTGTGCTGGTTTCCAGGGC 140899

Qy        999 AATGTCTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGT 1058
              |||||||  | | |||| |||||||||||||| ||| |||| |||| ||||||||||||||
Db     140900 AATGTCTCATTCCGAAGGTTCTAGAAACACAGTTTACTGGATGAAGAGCTGAAGTGGGGT 140959

Qy       1059 GGGG---GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATG 1115
              ||||   |         ||| ||||| |||||||||| |||||  |||||| ||| ||||
Db     140960 GGGGTAAGAATAACAACCACGTTCACTGAAGGGTGAGGGAGGACGGACAGACATACGATG 141019

Qy       1116 GCCAGCATAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
              ||||||||     ||| ||| |||||||||||  ||||||||| |||||||||| |
Db     141020 GCCAGCAT-----ACAAACATAACACCCTAATTAATGCTTCCCTCTGCTACTGACACTCT 141074

Qy       1176 CTTCACTCTCCTCTTTCATA------------------------AAAAATAAAAAAAGT 1210
              |    | ||| |||||||||                        ||||| |||| | ||
Db     141075 CCCTTCACTCTTCTTTCATAAAATAAACAAACAAACAAAAAAATCAAAAAAAAAATACGT 141134

Qy       1211 ATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATCTAAATATTTA 1270
              ||||||  || |||| | |||| ||||||| |     |||||||||| ||||||||||
Db     141135 ATTTTACATGACTCTGAAGATAAAATCTTTTC-----CTATAAAAAGACCTAAATATTTA 141189

Qy       1271 TATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATTTTTTGCCTCT 1330
              |||||||||  |||||| |||| || ||||| |||||||| ||||||| ||  |||||||
Db     141190 TATTTTTCATGTTTTAACATCTCAGTGATGATAAGCCAGAGACAAGTA-TTCCTGCCTCT 141248

Qy       1331 CTCAACAGCAAAGCTTGGGGCCTT-TTTGTTTCCGTGTTAGGAATAGAACACGAGAGCCC 1389
              |||||  |||||||||  |||||| ||||||| ||||||||||||||||
Db     141249 CTCAAGAGCAAAGCTTTTGGCCTTGTTTGTTTCGGTGTTAGGAATAGAACA--------- 141299

Qy       1390 CGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAGACGCACATT 1449
              | |||||||||||||||          ||| |||| ||||||||||||   ||| ||
Db     141300 CATGTATCTAGGCAGAT------------GCCTATGAATCTCCAGCCTCAAGTGCATGTT 141347

Qy       1450 TTTCTCGGGCTCTCTTAAGCTTT-TCCCACAGCATTGGGAAACTTTACTGACAGCATCCA 1508
              ||||||  |||||||  | |||||  |||||||||||||| || |||||||||||||| || |
Db     141348 TTTCTCAGGCTCTCCCAGGCTTTCCCCCACAGCATTGGAAAGCTTTACTGACAGCGTCAA 141407

Qy       1509 AGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGGCCCGTTTT 1568
              ||| |||||||||||| |||||| |||| || | ||| || | ||||  | ||| ||||
Db     141408 AGTCGTGCTTCTGCTAATAACTGGGCTCATATGTGTCTCTGGACCACTCTGACCCATTTT 141467

Qy       1569 GGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGTGCACTAAT 1628
              | |||  ||  || |||||||||| | |||||| |||         ||   |||||||||
Db     141468 GAGGTCAATATTCCGATTAGCCTTCTGTTTTAGAGCACA-------TGCCATGCACTAAT 141520

Qy       1629 TATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTTCTCCTAAT 1688
              |||||||||||||||||| || ||||||||||||||||||||||||||||||| ||||
Db     141521 TATGGCCAGTGACACCATAAAGTAAAAGTGCATTACTGAATGCTTTCAATTTCTCATAAT 141580

Qy       1689 GCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAGAGAATTCC 1748
              | ||||| | |||| |||||  ||||||||||||    ||||||||||||||||||||
Db     141581 GATGGTAAGGTGGCGTGTCATGGGGCCATTTTAGCCCTGGACATCACTCCAGAGAATTCT 141640

Qy       1749 AAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGATTATCCCC 1808
              |||||||| |||||||| |||| |||||||    ||||||| ||||||| |||| |||
Db     141641 GAACAGATATAGACAAGTGTCACCAAGACCCA----CTTCCCCCCGGGCTGTTTATTCCC 141696

Qy       1809 AGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTT 1868
              || ||||||||||||||||||||||||  ||||||||||||||||||| ||||||
Db     141697 AGGAATAGGATGTCCCAAAGCAACACTTCCAGGCCAACTGGAGTGCTGATAAGCCCAGTT 141756

Qy       1869 ATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGATACGTTAGT 1928
              |||||||||||| || || |||||||||||| |||||||| |||| ||||| ||||||
Db     141757 ATCAGAAAGATATTGCCGTGAGTGTGATGCACAATGCTTGCACTTTCCTGATAGGTTAGT 141816

Qy       1929 CATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTAACAGGCAG 1988
              |||| || |||||||| ||||||| |||| |||||||||||||||||||| |||||||||
Db     141817 CATACGAAAGCTGACAGAGAAGGAGAAAGGGCAGCGATGTGGTGCAATATGAACAGGCAG 141876

Qy       1989 CTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGC 2048
              |||| |||||||||||||||| || |||||||||||||||||| ||||||||||||||||
Db     141877 CTGT-CCCTGGCTTCCCGATAAGTAGGATGACTCGCATTGCTGGGCGGTGTGGTCACTGC 141935

Qy       2049 CAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCAT 2108
              ||| |||||| |||||||||||||||||||||||||| || || ||||||||||||||||
Db     141936 CAATGGAATGGCCCTCTCACATTTCTTCCTGATTCGCACACACCACGGCCAGCTTGTCAT 141995

Qy       2109 CTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTG 2168
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141996 CTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTG 142055

Qy       2169 GCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAG 2228
              |||||  | || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     142056 GCCACCAGCGGAGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAG 142115

Qy       2229 CCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCAT 2288
              |||||||||||||  |||||| ||||||||||||||||||||||| ||||||||||||
Db     142116 TCCTAGCTATATAAGAGGCTGGGGTGGAGGGGCTCCACAGGGCCAGCTCCAGGGGTTCAG 142175

Qy       2289 CCACAAGAGAGAAAAACATAGACTCGAGG 2317
              |||||||| ||||||||||||||||  ||
Db     142176 CCACAAGAGGGAAAAACATAGACTCACGG 142204


RESULT 4
AC097115
LOCUS       AC097115               238344 bp    DNA     linear   HTG 14-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-26A2, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC097115
VERSION     AC097115.6  GI:24956605
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238344)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
```

```
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238344)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238344)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 14, 2002 this sequence version replaced gi:22855482.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGQJ
                Center clone name: CH230-26A2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222569 bases at least Q40
                Consensus quality: 226638 bases at least Q30
                Consensus quality: 229427 bases at least Q20
                Estimated insert size: 233018; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   236546: contig of 236546 bp in length
            *  236547   236646: gap of unknown length
            *  236647   238344: contig of 1698 bp in length.
FEATURES            Location/Qualifiers
     source         1. .238344
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-26A2"
     misc_feature   1. .1081
                    /note="wgs_contig"
BASE COUNT    63454 a  47381 c  49225 g  71429 t   6855 others
ORIGIN

  Query Match             37.3%;  Score 879.8;  DB 2;  Length 238344;
  Best Local Similarity   82.3%;  Pred. No. 1e-245;
  Matches 1209;  Conservative    0;  Mismatches  177;  Indels   83;  Gaps   14;

Qy        879 ATGAAAACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
              ||| |||| ||||| |||||||| ||| | |||| |||||||||||  |||||||||
Db      26418 ATGGAAACCAGAAGGCCAACAGTTGTCCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 26477

Qy        939 AGAACACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGC 998
              |||||||||||| |||||||||||||     ||| |||||| |||||| ||||||
Db      26478 ----CACTAGGAGAGGAGAACCCACGAAGGA-----TATCAGTGTGCTGGTTTCCAGGGC 26528

Qy        999 AATGTCTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGT 1058
              |||||||  | | |||| |||||||||||||| ||| |||| |||| ||||||||||||||
Db      26529 AATGTCTCATTCCGAAGGTTCTAGAAACACAGTTTACTGGATGAAGAGCTGAAGTGGGGT 26588

Qy       1059 GGGG---GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATG 1115
              ||||   |      ||| |||||  |||||||||| |||||  |||||| ||| ||||
Db      26589 GGGGTAAGAATAACAACCACGTTCACTGAAGGGTGAGGGAGGACGGACAGACATACGATG 26648

Qy       1116 GCCAGCATAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
              ||||||||     ||| ||| ||||||||||||||  |||||||| |||||||||| |
Db      26649 GCCAGCAT-----ACAAACATAACACCCTAATTAATGCTTCCCTCTGCTACTGACACTCT 26703

Qy       1176 CTTCACTCTCCTCTTTCATA------------------------AAAAATAAAAAAAGT 1210
              |    | ||| |||||||||                        |||||  |||| | ||
Db      26704 CCCTTCACTCTTCTTTCATAAAATAAACAAACAAACAAAAAAATCAAAAAAAAAATACGT 26763

Qy       1211 ATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATCTAAATATTTA 1270
              ||||||  || |||| | |||| ||||||| |     ||||||||||| ||||||||||
Db      26764 ATTTTACATGACTCTGAAGATAAAATCTTTTC-----CTATAAAAAGACCTAAATATTTA 26818

Qy       1271 TATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATTTTTTGCCTCT 1330
              |||||||||  |||||| |||| || ||||| |||||||| ||||||| ||  |||||||
Db      26819 TATTTTTCATGTTTTAACATCTCAGTGATGATAAGCCAGAGACAAGTA-TTCCTGCCTCT 26877

Qy       1331 CTCAACAGCAAAGCTTGGGGCCTT-TTTGTTTCCGTGTTAGGAATAGAACACGAGAGCCC 1389
              ||||| |||||||||  |||||| |||||||| ||||||||||||||||||
Db      26878 CTCAAGAGCAAAGCTTTTGGCCTTGTTTGTTTCGGTGTTAGGAATAGAACA--------- 26928
```

```
Qy       1390 CGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAGACGCACATT 1449
             | ||||||||||||||||          ||| |||| ||||||||||||   |||  ||
Db      26929 CATGTATCTAGGCAGAT----------GCCTATGAATCTCCAGCCTCAAGTGCATGTT 26976

Qy       1450 TTTCTCGGGCTCTCTTAAGCTTT-TCCCACAGCATTGGGAAACTTTACTGACAGCATCCA 1508
             |||||| |||||||  | |||||  |||||||||||||| || |||||||||||||| || |
Db      26977 TTTCTCAGGCTCTCCCAGGCTTTCCCCCACAGCATTGGAAAGCTTTACTGACAGCGTCAA 27036

Qy       1509 AGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGGCCCGTTTT 1568
             ||| |||||||||||| |||||| |||| || | ||| || | ||||  | ||| ||||
Db      27037 AGTCGTGCTTCTGCTAATAACTGGGCTCATATGTGTCTCTGGACCACTCTGACCCATTTT 27096

Qy       1569 GGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGTGCACTAAT 1628
             | |||  ||  || ||||||||||| | |||||| |||         ||   |||||||||
Db      27097 GAGGTCAATATTCCGATTAGCCTTCTGTTTTAGAGCACA-------TGCCATGCACTAAT 27149

Qy       1629 TATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTTCTCCTAAT 1688
             |||||||||||||||||| ||| ||||||||||||||||||||||||||||||| ||||
Db      27150 TATGGCCAGTGACACCATAAAGTAAAAGTGCATTACTGAATGCTTTCAATTTCTCATAAT 27209

Qy       1689 GCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAGAGAATTCC 1748
             | ||||| | |||| |||||  |||||||||||||     ||||||||||||||||||
Db      27210 GATGGTAAGGTGGCGTGTCATGGGGCCATTTTAGCCCTGGACATCACTCCAGAGAATTCT 27269

Qy       1749 AAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGATTATCCCC 1808
              |||||||| |||||||| |||| |||||||     ||||||| ||||||| |||| |||
Db      27270 GAACAGATATAGACAAGTGTCACCAAGACCCA----CTTCCCCCCGGGCTGTTTATTCCC 27325

Qy       1809 AGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTT 1868
             || |||||||||||||||||||||||||  |||||||||||||||||||||| ||||||
Db      27326 AGGAATAGGATGTCCCAAAGCAACACTTCCAGGCCAACTGGAGTGCTGATAAGCCCAGTT 27385

Qy       1869 ATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGATACGTTAGT 1928
             |||||||||||| || || |||||||||||| |||||||| |||| ||||| ||||||
Db      27386 ATCAGAAAGATATTGCCGTGAGTGTGATGCACAATGCTTGCACTTTCCTGATAGGTTAGT 27445

Qy       1929 CATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTAACAGGCAG 1988
             |||| || |||||||| ||||||| |||| ||||||||||||||||||| ||||||||||
Db      27446 CATACGAAAGCTGACAGAGAAGGAGAAAGGGCAGCGATGTGGTGCAATATGAACAGGCAG 27505

Qy       1989 CTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGC 2048
             |||| |||||||||||||||| || |||||||||||||||||| ||||||||||||||||
Db      27506 CTGT-CCCTGGCTTCCCGATAAGTAGGATGACTCGCATTGCTGGGCGGTGTGGTCACTGC 27564

Qy       2049 CAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCAT 2108
             ||| |||||| |||||||||||||||||||||||||||||||| || || ||||||||||||
Db      27565 CAATGGAATGGCCCTCTCACATTTCTTCCTGATTCGCACACACCACGGCCAGCTTGTCAT 27624

Qy       2109 CTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTG 2168
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      27625 CTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTG 27684

Qy       2169 GCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAG 2228
             |||||  | || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      27685 GCCACCAGCGGAGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAG 27744

Qy       2229 CCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCAT 2288
              |||||||||||  |||||| |||||||||||||||||||| ||||||||||
Db      27745 TCCTAGCTATATAAGAGGCTGGGGTGGAGGGGCTCCACAGGGCCAGCTCCAGGGGTTCAG 27804

Qy       2289 CCACAAGAGAGAAAAACATAGACTCGAGG 2317
             ||||||||| ||||||||||||||  ||
Db      27805 CCACAAGAGGGAAAAACATAGACTCACGG 27833


RESULT 5
BV062893/c
LOCUS       BV062893                838 bp    DNA     linear   STS 31-MAY-2003
DEFINITION  S212P60252FC3.T0 CZECHII/Ei Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV062893
VERSION     BV062893.1  GI:31178688
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
  TITLE     The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
  MEDLINE   22354684
   PUBMED   12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 838
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="CZECHII/Ei"
                     /db_xref="taxon:10090"
                     /map="+ 19 27-787 35750684-35751444"
                     /clone_lib="CZECHII/Ei"
     STS             <1. .>838
BASE COUNT      235 a    164 c    208 g    231 t
ORIGIN

  Query Match           34.1%;  Score 804.8;  DB 11;  Length 838;
  Best Local Similarity   98.9%;  Pred. No. 5.2e-224;
  Matches  831;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy       1004 CTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGG 1063
              || || |||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db        838 CTCGTTCTGAAGGTTCTAGAAACACAATTTGCTGG-TGAACAGCTGAAGTGGGGTGGGGG 780

Qy       1064 TTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCAT 1123
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        779 TTCTTACCCCATGTTCATTGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCAT 720

Qy       1124 AACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTC 1183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 AACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTC 660

Qy       1184 TCCTCTTTCATAAAAAATAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCT 1243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TCCTCTTTCATAAAAAATAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCT 600

Qy       1244 CGAACTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACA 1303
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        599 CGAACTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAATATCTTAGGGATGACA 540

Qy       1304 AGCCAGAAACAAGTATTTTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCC 1363
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGCCAGAAACAAGTA-TTTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCC 481

Qy       1364 GTGTTAGGAATAGAACACGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCA 1423
```

```
Db        480 GTGTTAGGAATAGAACACGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCA 421

Qy       1424 TGAGTCTCCAGCCTCAGACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCAT 1483

Db        420 TGAGTCTCCAGCCTCAGACGCACATTTTTCTCGGGCTCTCTTAAGCTTTCCCCACAGCAT 361

Qy       1484 TGGGAAACTTTACTGACAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTC 1543

Db        360 TGGGAAACTTTACTGACAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTC 301

Qy       1544 TCTGTGCATCACTTCGGCCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAA 1603

Db        300 TCTGTGCATCACTTCGGCCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAA 241

Qy       1604 CACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTA 1663

Db        240 CACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTA 181

Qy       1664 CTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGC 1723

Db        180 CTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACGGGGCCATTTTAGC 121

Qy       1724 TGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCT 1783

Db        120 TGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCT 61

Qy       1784 CCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCC 1843

Db          60 CCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCC 1
```

```
RESULT 6
AF478692
LOCUS       AF478692                723 bp    DNA     linear   ROD 10-OCT-2002
DEFINITION  Mus musculus cardiac ankyrin repeat protein (Carp) gene, promoter
            region and partial cds.
ACCESSION   AF478692
VERSION     AF478692.1  GI:19110906
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Maeda,T., Sepulveda,J., Chen,H.H. and Stewart,A.F.R.
  TITLE     alpha1-Adrenergic activation of the cardiac ankyrin repeat protein
            gene in cardiac myocytes
  JOURNAL   Gene 297 (1-2), 1-9 (2002)
REFERENCE   2  (bases 1 to 723)
  AUTHORS   Maeda,T., Sepulveda,J. and Stewart,A.F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Cardiovascular Institute, University of
            Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15213, USA
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="FVB"
                     /db_xref="taxon:10090"
     gene            1. .>723
                     /gene="Carp"
     misc_feature    1. .720
                     /gene="Carp"
                     /note="contains promoter and 5' UTR"
     mRNA            <721. .>723
                     /gene="Carp"
                     /product="cardiac ankyrin repeat protein"
     CDS             721. .>723
                     /gene="Carp"
                     /codon_start=1
                     /product="cardiac ankyrin repeat protein"
                     /protein_id="AAL85342.1"
                     /db_xref="GI:19110907"
                     /translation="M"
BASE COUNT      187 a    185 c    180 g    171 t
ORIGIN
```

```
  Query Match           27.4%;  Score 646.2;  DB 10;  Length 723;
  Best Local Similarity 97.5%;  Pred. No. 1.5e-177;
  Matches  699;  Conservative    0;  Mismatches   13;  Indels    5;  Gaps    4;

Qy       1601 GAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCA 1660

Db          1 GAACACGGTGAGCCTCTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTCGA 60

Qy       1661 TTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTT 1720

Db         61 TTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTT 120

Qy       1721 AGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCA 1780

Db        121 AGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCA 180

Qy       1781 TCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCA 1840

Db        181 TCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATA-GATGTCCCAAAGCA--ACTTCCCA 237

Qy       1841 GCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCAC 1900

Db        238 GCCAACTGGATGCGTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATCGAC 297

Qy       1901 AGTGCTTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGC 1960

Db        298 AGTGCTTCGATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGG-AAAAGAGC 356

Qy       1961 AGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGAC 2020

Db        357 AGCGATGT-GTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGAC 415

Qy       2021 TCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGA 2080

Db        416 TCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGA 475

Qy       2081 TTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTG 2140

Db        476 TTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTG 535

Qy       2141 GAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCT 2200

Db        536 GAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCT 595

Qy       2201 TCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGG 2260

Db        596 TCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGG 655

Qy       2261 CTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGG 2317

Db        656 CTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAGACTCACGG 712
```

```
RESULT 7
BV076484
LOCUS       BV076484                665 bp    DNA     linear   STS 31-MAY-2003
DEFINITION  S212P60369FF7.T0 CZECHII/Ei Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV076484
VERSION     BV076484.1  GI:31192279
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
```

```
  TITLE     The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
  MEDLINE   22354684
   PUBMED   12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 665
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
     source          1. .665
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="CZECHII/Ei"
                     /db_xref="taxon:10090"
                     /map="- 19 22-610 35751955-35751368"
                     /clone_lib="CZECHII/Ei"
     STS             <1. .>665
BASE COUNT     191 a    155 c    164 g    155 t
ORIGIN

  Query Match           27.2%;  Score 640.2;  DB 11;  Length 665;
  Best Local Similarity   98.6%;  Pred. No. 8.3e-176;
  Matches  656;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        524 GTAGGAACTGTATGTACCAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTG 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTAGGAACTGTATGTACCAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTG 60

Qy        584 GAGCGAAGACGCAACCTCACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGCGAAGACGCAACCTCACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCT 120

Qy        644 GTGTCTAGGTGTCATCTCTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCAC 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGTCTAGGTGTCATCTCTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCAC 180

Qy        704 AAAATACTCAGAGATTCAAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTC 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAATACTCAGAGATTCAAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTC 240

Qy        764 TCAACTCTCAGCCCCCCTTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAG 823
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCAACTCTCAGCCCCACTTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAG 300

Qy        824 AAAAGCTGCCATAGCTGCTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAA 883
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGCTGCCATAGCTGCTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAA 360

Qy        884 AACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAAC 943
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        361 AACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATGGGACAGAGAGAAC 420

Qy        944 ACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGT 1003
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        421 ACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAACGT 480

Qy       1004 CTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGT-GGGGTGGGG 1062
              || || |||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db        481 CTCGTTCTGAAGGTTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGGTGGGG 540

Qy       1063 GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCA 1122
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        541 GTTCTTACCCCATGTTCATTGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCA 600

Qy       1123 TAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACT 1182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACT 660

Qy       1183 CTCCT 1187
              |||||
Db        661 CTCCT 665


RESULT 8
AX468604
LOCUS       AX468604               2074 bp    DNA     linear   PAT 16-JUL-2002
DEFINITION  Sequence 2 from Patent WO0246220.
ACCESSION   AX468604
VERSION     AX468604.1  GI:21901403
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Schwartz,B., Branellec,D. and Chien,K.
  TITLE     Sequences upstream of the carp gene, vectors containing them and
            uses thereof
  JOURNAL   Patent: WO 0246220-A 2 13-JUN-2002;
            Aventis Pharma S.A. (FR) ; The Regents of The University of
            California at San Diego (US) ; Benoit, Patrick (FR)
FEATURES             Location/Qualifiers
     source          1. .2074
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      612 a    469 c    416 g    572 t      5 others
ORIGIN

  Query Match           15.6%;  Score 367.8;  DB 6;  Length 2074;
  Best Local Similarity   76.4%;  Pred. No. 5.9e-96;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
              ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||  | | | |
Db       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
              |||||||||||| ||| |||  ||  |||   |||||||||||||| |||| |   |||||
Db       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
              |||||||||||||||||||| |||  ||||| ||||| | ||||||||||  |||||
Db       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465

Qy       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
              ||||||  |||||||||||||| |||||||||||||||| ||||||||  |     | |
Db       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524

Qy       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837
              |||  ||| | |||     |    ||  |||  | ||||||||||||      |||   | |
Db       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGGACAAGTTTCC 1584

Qy       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897
              |   | || || |||| |||||||||   ||||||||||||||  |||   |||||||
Db       1585 C---CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641

Qy       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955
                 || ||   | |||||||||||| | |||||||||| ||||||||||||  ||||
Db       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698
```

```
Qy       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015
             || |||| |||||||||||||| | ||||| |||||||||||||| | ||  || |  | ||
Db       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757

Qy       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075
             |||||| |||||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817

Qy       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130
             |||||||| ||||  | || |      |  |  ||  ||||||||  ||||||||||
Db       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877

Qy       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190
             ||| |||||||||||||||||||||||||||||| |||||  ||  |||||||  || |||
Db       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937

Qy       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249
             | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997

Qy       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309
             |||||||||||| |  |||||||| ||||||| ||| | | || || ||||||||||
Db       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056

Qy       2310 A 2310
             |
Db       2057 A 2057
```

```
RESULT 9
AF131884
LOCUS       AF131884                2074 bp    DNA     linear   PRI 09-FEB-2000
DEFINITION  Homo sapiens cardiovascular-specific cardiac ankyrin repeat protein
            (CVARP) gene, 5'-flanking region and exon 1.
ACCESSION   AF131884
VERSION     AF131884.1  GI:6940841
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2074)
  AUTHORS   Aihara,Y., Kurabayashi,M., Tanaka,T., Sekiguchi,K., Tomaru,K.,
            Kanai,H., Takeda,S. and Nagai,R.
  TITLE     Human CVARP 5'-flanking region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2074)
  AUTHORS   Aihara,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1999) Second Department of Internal Medicine,
            Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi
            371-8511, Japan
FEATURES             Location/Qualifiers
     source          1. .2074
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1. .1832
                     /note="5'-flanking region"
     gene            1833. .>2074
                     /gene="CVARP"
     mRNA            1833. .>2074
                     /gene="CVARP"
                     /product="cardiovascular-specific cardiac ankyrin repeat
                     protein"
                     /note="CARP"
     exon            1833. .2074
                     /gene="CVARP"
                     /number=1
BASE COUNT      612 a    469 c    416 g    572 t      5 others
ORIGIN
```

```
  Query Match           15.6%;  Score 367.8;  DB 9;  Length 2074;
  Best Local Similarity  76.4%;  Pred. No. 5.9e-96;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
             ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
             |||||||||||| ||| |||  ||  |||  |||||||||||||| |||| |   |||||
Db       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
             |||||||||||||||||||| |||  ||||| ||||| | ||||||||||  |||||
Db       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465

Qy       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
             ||||||  |||||||||||||| |||||||||||||||| ||||||||| |     | ||
Db       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524

Qy       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837
             |||  ||| | |||     |    ||  |||  | ||||||||||||      |||   | |
Db       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGGACAAGTTTCC 1584

Qy       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897
             |   | || || |||| |||||||||   ||||||||||||||||  |||   ||||||||
Db       1585 C---CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641

Qy       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955
                 ||  ||    | |||||||||||||| | |||||||||| ||||||||||||  ||||
Db       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698

Qy       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015
             || |||| |||||||||||||| | ||||| |||||||||||||| | ||  || |  | ||
Db       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757

Qy       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075
             |||||| |||||||||||||| |||| |||  |||||||||| ||||||||||||||||
Db       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817

Qy       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130
             |||||||| ||||  | || |      |  |  ||  ||||||||  ||||||||||
Db       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877

Qy       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190
             ||| |||||||||||||||||||||||||||||| |||||  ||  |||||||  || |||
Db       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937

Qy       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249
             | |||| | |||||||| |||||||||||||||||| |||| |||||||||| | | | |
Db       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997

Qy       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309
             |||||||||||| |  |||||||| ||||||| ||| | | || || ||||||||||
Db       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056

Qy       2310 A 2310
             |
Db       2057 A 2057
```

```
RESULT 10
AL590622/c
LOCUS       AL590622                50111 bp    DNA     linear   PRI 19-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-320F15 on chromosome 10.
            Contains the gene for ribonuclease P 30kD subunit, the gene for a
            nuclear protein similar to CARP, ESTs, STSs and GSSs, complete
            sequence.
ACCESSION   AL590622
VERSION     AL590622.7  GI:14270159
KEYWORDS    HTG; CARP; ribonuclease P.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50111)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 31, 2001 this sequence version replaced gi:14161205.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-320F15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-320F15 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-236B18 is at 50012 in this
            sequence. The true right end of clone RP11-103A2 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .50111
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-320F15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(1. .98)
                     /note="match: GSS: Em:AQ670367"
     misc_feature    complement(1. .97)
                     /note="match: GSS: Em:AQ544636"
     misc_feature    complement(1. .77)
                     /note="match: STS: Em:G56439
                     match: GSS: Em:AQ311708"
     gene            120. .34469
                     /gene="bA320F15.1"
     mRNA            join(120. .472,3273. .3328,3962. .4018,4428. .4502,7467. .7538,
                     14216. .14305,23179. .23295,23858. .23887,24284. .24321,
                     24717. .24796,28974. .29381)
                     /gene="bA320F15.1"
                     /product="bA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
                     /note="match: cDNAs: Em:U77665 Em:AK004137 Em:U95123
                     match: ESTs: Em:BE572689 Em:BE378859 Em:BE137893
                     Em:AA920575 Em:BG106820 Em:BF681617 Em:BF248001
                     Em:BF031745 Em:AA854455 Em:AI359795"
                     /evidence=not_experimental
     CDS             join(391. .472,3273. .3328,3962. .4018,4428. .4502,7467. .7538,
                     14216. .14305,23179. .23295,23858. .23887,24284. .24321,
                     24717. .24796,28974. .29083)
                     /gene="bA320F15.1"
                     /note="match: proteins: Tr:P78346 Tr:O88796"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
                     /protein_id="CAC70100.1"
                     /db_xref="GI:15717970"
                     /translation="MAVFADLDLRAGSDLKALRGLVETAAHLGYSVVAINHIVDFKEK
                     KQEIEKPVAVSELFTTLPIVQGKSRPIKILTRLTIIVSDPSHCNVLRATSSRARLYDV
                     VAVFPKTEKLFHIACTHLDVDLVCITVTEKLPFYFKRPPINVAIDRGLAFELVYSPAI
                     KDSTMRRYTISSALNLMQICKGKNVIISSAAERPLEIRGPYDVANLGLLFGLSESDAK
                     AAVSTNCRAALLHGETRKTAFGIISTVKKPRPSEGDEDCLPASKKAKCEG"
     repeat_region   689. .800
                     /note="MIR repeat: matches 30. .132 of consensus"
     repeat_region   1443. .1488
                     /note="L2 repeat: matches 2459. .2502 of consensus"
     misc_feature    1761. .2155
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ805524"
     misc_feature    1907. .2434
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ775672"
     misc_feature    1926. .2208
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:B92129"
     repeat_region   2349. .2482
                     /note="L2 repeat: matches 2557. .2696 of consensus"
     repeat_region   2939. .3167
                     /note="MIR repeat: matches 13. .262 of consensus"
     misc_feature    complement(4176. .4649)
                     /note="match: GSS: Em:AQ626500"
     mRNA            join(<4648. .4764,7467. .7538,14216. .14305,23179. .23887,
                     24284. .24321,24717. .24796,28974. .29381)
                     /gene="bA320F15.1"
                     /product="bA320F15.1.2 (putative isoform 2)"
                     /note="match: ESTs: Em:AW939965"
                     /evidence=not_experimental
     repeat_region   4733. .4858
                     /note="L2 repeat: matches 2601. .2732 of consensus"
     repeat_region   4928. .5102
                     /note="L2 repeat: matches 2299. .2478 of consensus"
     repeat_region   5257. .5439
                     /note="L2 repeat: matches 1808. .1971 of consensus"
     repeat_region   5440. .5721
                     /note="AluSx repeat: matches 7. .293 of consensus"
     repeat_region   5722. .6237
                     /note="L2 repeat: matches 1200. .1808 of consensus"
     repeat_region   6884. .7196
                     /note="AluJb repeat: matches 1. .312 of consensus"
     repeat_region   8243. .8404
                     /note="2 copies 81 mer 81% conserved"
     repeat_region   8445. .8541
                     /note="HY1 repeat: matches 16. .112 of consensus"
     repeat_region   8542. .8583
                     /note="HY1 repeat: matches 1. .42 of consensus"
     repeat_region   8850. .9130
                     /note="AluSx repeat: matches 1. .289 of consensus"
     repeat_region   9601. .9741
                     /note="MIR repeat: matches 76. .212 of consensus"
     repeat_region   10798. .11076
                     /note="AluY repeat: matches 1. .291 of consensus"
     repeat_region   11086. .11219
                     /note="FLAM_C repeat: matches 1. .129 of consensus"
     repeat_region   11319. .11491
                     /note="MER5A repeat: matches 3. .189 of consensus"
     repeat_region   11928. .11984
                     /note="L2 repeat: matches 1743. .1801 of consensus"
     repeat_region   13755. .13817
                     /note="L2 repeat: matches 2684. .2749 of consensus"
     repeat_region   15390. .15671
                     /note="AluJo repeat: matches 1. .282 of consensus"
     repeat_region   15791. .16088
                     /note="AluSg repeat: matches 1. .296 of consensus"
     repeat_region   16979. .17036
                     /note="L2 repeat: matches 2423. .2481 of consensus"
```

```
     misc_feature    17236. .17755
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ551877"
     repeat_region   17709. .17972
                     /note="AluSx repeat: matches 45. .304 of consensus"
     repeat_region   19198. .19363
                     /note="L1ME repeat: matches 5272. .5435 of consensus"
     repeat_region   19489. .19928
                     /note="MLT1C repeat: matches 5. .478 of consensus"
     repeat_region   21838. .22258
                     /note="L2 repeat: matches 2109. .2669 of consensus"
     misc_feature    22440. .22946
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ876596"
     mRNA            join(23179. .23295,23858. .23887,24284. .24321,24717. .24796,
                     28974. .29072,30647. .30748,31596. .31680,33706. .34469)
                     /gene="bA320F15.1"
                     /product="bA320F15.1.3 (putative isoform 3)"
                     /note="match: ESTs: Em:BE567341 Em:AI292002 Em:BF684192
                     Em:BF435407 Em:AI740881 Em:AA651912 Em:BF438915"
                     /evidence=not_experimental
     misc_feature    24121. .24518
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ026750"
     misc_feature    complement(join(24284. .24321,24724. .24796,28974. .29240))
                     /note="match: STS: Em:G24365"
     misc_feature    24549. .25020
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:AQ223714"
     misc_feature    24561. .24956
                     /gene="bA320F15.1"
                     /note="match: GSS: Em:B53566"
     repeat_region   25321. .25458
                     /note="FLAM_C repeat: matches 1. .133 of consensus"
     misc_feature    complement(25417. .25863)
                     /note="match: GSS: Em:AQ712354"
     repeat_region   26204. .26603
                     /note="L1MA2 repeat: matches 5888. .6302 of consensus"
     repeat_region   28442. .28516
                     /note="L2 repeat: matches 2636. .2710 of consensus"
     polyA_signal    29149. .29154
                     /gene="bA320F15.1"
     polyA_site      29166
                     /gene="bA320F15.1"

  Query Match             14.7%;  Score 347.6;  DB 9;  Length 50111;
  Best Local Similarity   78.2%;  Pred. No. 7e-90;
  Matches  521;  Conservative    0;  Mismatches  129;  Indels   16;  Gaps    8;

Qy       1653 AAAGTGCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGG 1712
             ||||||||||||||||||||||||||||||  |||||  |||||  | ||||||||||  ||
Db      50106 AAAGTGCATTACTGAATGCTTTCAATTTCTTATAATGATGGTAAGGTGGCATGTCATGGG 50047

Qy       1713 GCCATTTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACC 1772
              |||  ||||||  |||||||||||||| ||||||||||||||||  ||||||||| |
Db      50046 GCCTATTTAGCCCCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTT 49987

Qy       1773 CAGACCCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAAC 1832
                | ||||  |||| | |||     |    ||  |||  | ||||||||||||     |  ||
Db      49986 AGGGCCCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGAC 49930

Qy       1833 ACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTG 1892
              |  |  |  | || || |||| |||||||||   |||||||||||||||  |||   |||
Db      49929 AAGTTTCCCCTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTG 49870

Qy       1893 TGATGCACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAG 1950
              ||||    || ||   | |||||||||||| | |||||||||| ||||||||||||
Db      49869 TGATATGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA- 49812

Qy       1951 GAAAAAGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATAC 2010
              ||||||  |||| |||||||||||| | ||||| |||||||||||| | ||  || | |
Db      49811 -AAAAAGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAA 49754

Qy       2011 GTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACAT 2070
              | ||||||||  |||||||||||||  |||| ||||  ||||||||||  ||||||||||
Db      49753 ATAGGATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACAT 49694

Qy       2071 TTCTTCCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCC 2125
              ||||||||||||| ||||  | || |       |   |  ||  |||||||||  ||||||
Db      49693 TTCTTCCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCC 49634

Qy       2126 CAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAG 2185
              |||||||| |||||||||||||||||||||||||||||||||||||| |||||  ||  ||||||  |
Db      49633 CAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGTGG 49574

Qy       2186 GGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CG 2244
              | |||| |||| | |||||||| |||||||||||||||||| |||| |||||||||| |
Db      49573 GAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCT 49514

Qy       2245 GGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAA 2304
              | | |||||||||||||| |  |||||||  |||||||  ||| | | || || |||||||
Db      49513 GACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAA 49455

Qy       2305 CATAGA 2310
              |||| |
Db      49454 CATACA 49449


RESULT 11
AX322775
LOCUS       AX322775               1901 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 19 from Patent WO0192567.
ACCESSION   AX322775
VERSION     AX322775.1  GI:18093755
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Bunk,D., Reuner,B., Beck,J. and Henkel,T.
  TITLE     Novel target genes for diseases of the heart
  JOURNAL   Patent: WO 0192567-A 19 06-DEC-2001;
            Medigene AG (DE)
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match             5.5%;  Score 128.6;  DB 6;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 6.9e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  |||||||||||||| | ||||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
              |||||  ||  ||||||  || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | |||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | | || || |||||||||||  |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 12
BD094076
LOCUS       BD094076               1901 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Shear stress-responsive DNAs.
```

```
ACCESSION   BD094076
VERSION     BD094076.1  GI:22639664
KEYWORDS    WO 0125427-A/37.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1901)
  AUTHORS   Nojima,H., Yoshisue,H., Obayashi,M., Ota,T., Kawabata,A.,
            Sakurada,K., Kuga,T., Sekine,S., Nakamura,Y. and Sugan,S.
  TITLE     Shear stress-responsive DNAs
  JOURNAL   Patent: WO 0125427-A 37 12-APR-2001;
            KYOWA HAKKO KOGYO CO LTD,HIROSHI NOJIMA,HAJIME YOSHISUE, MASAYA
            OBAYASHI, TOSHIO OTA,AYAKO KAWABATA,KAZUHIRO SAKURADA,TETSURO KUGA,
            SUSUMU SEKINE, YUSUKE NAKAMURA,SUMIO SUGANO
COMMENT     OS   Homo sapiens (human)
            PN   WO 0125427-A/37
            PD   12-APR-2001
            PF   02-OCT-2000 WO 2000JP006840
            PR   01-OCT-1999 JP 99P   280976
            PI   HIROSHI NOJIMA,HAJIME YOSHISUE,MASAYA OBAYASHI,TOSHIO OTA, PI
             AYAKO KAWABATA,
            PI   KAZUHIRO SAKURADA,TETSURO KUGA,SUSUMU SEKINE,YUSUKE NAKAMURA,
            PI   SUMIO SUGANO
            PC   C12N15/12,C07K14/435,C07K16/18,C12P21/02,C12Q1/68,A61K38/00,
            PC   A61K39/395,
            PC   A61K48/00,A61P9/10,G01N33/50,G01N33/53
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (250). .(1206).
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match          5.5%;  Score 128.6;  DB 6;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 6.9e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  |||||||||||||| | ||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
               |||||  ||  ||||||  || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | |||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | |. || || |||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 13
HSRNACINP
LOCUS       HSRNACINP               1901 bp    mRNA    linear   PRI 05-SEP-1995
DEFINITION  H.sapiens mRNA for cytokine inducible nuclear protein.
ACCESSION   X83703
VERSION     X83703.1  GI:793840
KEYWORDS    ankyrin-like repeat; nuclear localisation signal; nuclear protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chu,W., Burns,D.K., Swerlick,R.A. and Presky,D.H.
  TITLE     Identification and characterization of a novel cytokine-inducible
            nuclear protein from human endothelial cells
  JOURNAL   J. Biol. Chem. 270 (17), 10236-10245 (1995)
  MEDLINE   95247734
   PUBMED   7730328
REFERENCE   2  (bases 1 to 1901)
  AUTHORS   Chu,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1995) W. Chu, Hoffmann-La Roche, 340 Kingsland
            Street, Dept. of Inflammation/Autoimmune Disease, Hoffmann-La
            Roche, Nutley, NJ 07110, USA
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="C-193"
                     /cell_type="endothelial"
                     /tissue_type="skin"
                     /clone_lib="HDMEC cDNA"
     mRNA            1. .1901
     misc_feature    94. .98
                     /note="nuclear localization signal"
     repeat_unit     152. .283
                     /note="ankyrin-like repeats"
     CDS             250. .1209
                     /note="cytokine-inducible expression"
                     /codon_start=1
                     /product="nuclear protein"
                     /protein_id="CAA58676.1"
                     /db_xref="GI:793841"
                     /db_xref="SPTREMBL:Q15327"
                     /translation="MMVLKVEELVTGKKNGNGEAGEFLPEDFRDGEYEAAVTLEKQED
                     LKTLLAHPVTLGEQQWKSEKQREAELPKKKLEQRSKLENLEDLEIIIQLKKRKKYRKT
                     KVPVVKEPEPEIITEPVDVPTFLKAALENKLPVVEKFLSDKNNPDVCDEYKRTALHRA
                     CLEGHLAIVEKLMEAGAQIEFRDMLESTAIHWASRGGNLDVLKLLLNKGAKISARDKL
                     LSTALHVAVRTGHYECAEHLIACEADLNAKDREGDTPLHDAVRLNRYKMIRLLIMYGA
                     DLNIKNCAGKTPMDLVLHWQNGTKAIFDSLRENSYKTSRIATF"
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match          5.5%;  Score 128.6;  DB 9;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 6.9e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  |||||||||||||| | ||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
               |||||  ||  ||||||  || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | |||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | |. || || |||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 14
G28603
LOCUS       G28603                  1901 bp    DNA     linear   STS 11-JUL-1996
DEFINITION  human STS SHGC-35401, sequence tagged site.
ACCESSION   G28603
VERSION     G28603.1  GI:1408418
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1901)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: GGCATTTTGAAGGCATGG
            Primer B: CCAGATGGATGATCATGAAGG
            STS size: 222
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from X83703
            -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="10"
     STS             1246. .1467
     primer_bind     1246. .1263
     primer_bind     complement(1447. .1467)
BASE COUNT      592 a    378 c    460 g    471 t
ORIGIN

  Query Match             5.5%;  Score 128.6;  DB 11;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 6.9e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  |||||||||||||| | ||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
              |||||  ||  ||||||  || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | |||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | | || || |||||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 15
AC074094/c
```

```
LOCUS       AC074094              160350 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 10 clone RP11-236B18, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC074094
VERSION     AC074094.3  GI:9958197
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160350)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160350)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:9665205.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0236B18
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 149545 bases at least Q40
            Consensus quality: 152532 bases at least Q30
            Consensus quality: 153924 bases at least Q20
            Insert size: 151000; agarose-fp
            Insert size: 158750; sum-of-contigs
            Quality coverage: 4.85 in Q20 bases; agarose-fp
            Quality coverage: 4.71 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      2326: contig of 2326 bp in length
            *     2327      2426: gap of unknown length
            *     2427      7519: contig of 5093 bp in length
            *     7520      7619: gap of unknown length
            *     7620     15286: contig of 7667 bp in length
            *    15287     15386: gap of unknown length
            *    15387     23050: contig of 7664 bp in length
            *    23051     23150: gap of unknown length
            *    23151     33663: contig of 10513 bp in length
            *    33664     33763: gap of unknown length
            *    33764     50133: contig of 16370 bp in length
            *    50134     50233: gap of unknown length
            *    50234     68088: contig of 17855 bp in length
            *    68089     68188: gap of unknown length
            *    68189     88133: contig of 19945 bp in length
            *    88134     88233: gap of unknown length
            *    88234    113397: contig of 25164 bp in length
            *   113398    113497: gap of unknown length
            *   113498    141732: contig of 28235 bp in length
            *   141733    141832: gap of unknown length
            *   141833    143187: contig of 1355 bp in length
            *   143188    143287: gap of unknown length
            *   143288    144942: contig of 1655 bp in length
```

```
            *   144943   145042: gap of unknown length
            *   145043   147286: contig of 2244 bp in length
            *   147287   147386: gap of unknown length
            *   147387   150852: contig of 3466 bp in length
            *   150853   150952: gap of unknown length
            *   150953   153550: contig of 2598 bp in length
            *   153551   153650: gap of unknown length
            *   153651   157257: contig of 3607 bp in length
            *   157258   157357: gap of unknown length
            *   157358   160350: contig of 2993 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160350
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-236B18"
     misc_feature    1. .2326
                     /note="assembly_name:Contig10"
     misc_feature    2427. .7519
                     /note="assembly_name:Contig11"
     misc_feature    7620. .15286
                     /note="assembly_name:Contig12"
     misc_feature    15387. .23050
                     /note="assembly_name:Contig13"
     misc_feature    23151. .33663
                     /note="assembly_name:Contig14"
     misc_feature    33764. .50133
                     /note="assembly_name:Contig15"
     misc_feature    50234. .68088
                     /note="assembly_name:Contig16"
     misc_feature    68189. .88133
                     /note="assembly_name:Contig17"
     misc_feature    88234. .113397
                     /note="assembly_name:Contig18"
     misc_feature    113498. .141732
                     /note="assembly_name:Contig19"
     misc_feature    141833. .143187
                     /note="assembly_name:Contig3"
     misc_feature    143288. .144942
                     /note="assembly_name:Contig4"
     misc_feature    145043. .147286
                     /note="assembly_name:Contig5"
     misc_feature    147387. .150852
                     /note="assembly_name:Contig6"
     misc_feature    150953. .153550
                     /note="assembly_name:Contig7"
     misc_feature    153651. .157257
                     /note="assembly_name:Contig8"
     misc_feature    157358. .160350
                     /note="assembly_name:Contig9"
BASE COUNT    47390 a  32724 c  33135 g  45498 t   1603 others
ORIGIN

  Query Match             5.4%;  Score 127.2;  DB 2;  Length 160350;
  Best Local Similarity   52.5%;  Pred. No. 2.9e-25;
  Matches  171;  Conservative    0;  Mismatches  153;  Indels    2;  Gaps    1;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
              ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db     113707 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 113648

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
              |||||||||||| ||| |||  ||  |||||||||||||||||||| |||| |   |||||
Db     113647 TTAGAACACGGTAAGCATGTCATGTGCTAATTATGGCCAGTGACATCATAAAAGAAAAGT 113588

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
              |||||||||||||||||||||||  ||||| ||||| | ||||||||||  |||||
Db     113587 GCATTACTGAATGCTTTCAATTTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 113528

Qy       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
              ||||||  |||||||||||||| |||||||
Db     113527 TTTAGCCCCAGACATCACTCCAAAGAATTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113468

Qy       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837

Db     113467 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113408

Qy       1838 CCAGCCAACTGGAGTGCTGATAAGTC 1863
                        | | | |+| ||  ||
Db     113407 NNNNNNNNNNTGTGAGATGGTATCTC 113382


Search completed: November 18, 2003, 04:21:58
Job time : 8596.58 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:28:30 ; Search time 618.23 Seconds
(without alignments)
10295.963 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggatcctttcatgtttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2358	100.0	2358	24	ABV73020	Mouse CARP protein
2	1763.4	74.8	2247	21	AAA10406	Murine cardiac ank
3	367.8	15.6	2074	24	ABV73021	Human CARP protein
4	130.2	5.5	1988	22	AAI93508	Human polynucleoti
5	128.6	5.5	1901	22	AAH02910	Human shear stress
6	128.6	5.5	1901	24	AAD27217	Human CAA58676 pro
7	95	4.0	7280	25	ABZ24607	Mouse gastric antr
8	95	4.0	7280	25	AAD50815	Mouse pre-AMP-18 g
c 9	89.2	3.8	18105	25	ABZ82337	Mouse histamine re
10	88.8	3.8	659158	25	ABX16390	Mouse high growth
c 11	88.6	3.8	48974	20	AAX55300	Mouse Presenilin-1
12	83.6	3.5	707	24	ABT09109	Phase-1 Rat CT gen
c 13	82	3.5	659158	25	ABX16390	Mouse high growth
14	80.4	3.4	10917	24	ABA97708	Wild-type mouse Wo
15	78	3.3	892	18	AAV06155	Viral infection ge
c 16	76.8	3.3	249487	24	ABN85733	Mouse genomic regi
17	76.6	3.2	5023	24	ABL56461	Nucleotide sequenc
18	76.6	3.2	17758	24	ABL56463	ALpha6 locus of al
19	75.8	3.2	335	24	ABS69028	Novel murine polyn
c 20	75.6	3.2	5990	24	ABK62570	Rat sequence diffe
c 21	75.6	3.2	7280	25	ABZ24607	Mouse gastric antr
c 22	75.6	3.2	7280	25	AAD50815	Mouse pre-AMP-18 g
c 23	75	3.2	49999	20	AAZ23891	Murine LOBO genomi
c 24	75	3.2	49999	20	AAZ23896	Murine LOBO homolo
25	74.8	3.2	37940	20	AAZ01026	Partial mouse PG1
c 26	74.6	3.2	10614	16	AAQ89555	Hamster cholestero
27	73.4	3.1	4164	19	AAV12216	Mouse retinoid met
28	73.4	3.1	4164	24	AAD24512	Mouse P450RAI geno
c 29	71.6	3.0	171936	24	ABS56565	Human SULF2 genomi
30	69.4	2.9	1889	24	AAS94903	Human DNA sequence
31	69.4	2.9	1889	25	ABX77648	Differentially exp
32	68.6	2.9	4072	22	AAC85818	Murine sequence fr
33	68	2.9	1094	20	AAX57479	Rat U3 gene trap d
c 34	68	2.9	16442	18	AAX83006	Partial mouse WRN
35	65.6	2.8	29604	18	AAX83005	Partial mouse WRN
c 36	65	2.8	6645	17	AAT65002	Mouse cell cycle r
37	63	2.7	874	20	AAX57448	Rat U3 gene trap d
38	63	2.7	874	20	AAX57470	Rat U3 gene trap d
39	63	2.7	3240	22	AAS05314	Mouse alpha-1,3 ga
40	63	2.7	3537	22	AAS05315	Mouse alpha-1,3 ga
c 41	62	2.6	7208	21	AAA40866	Murine tumour necr
42	61.6	2.6	8402	24	ABK87055	Genomic DNA encodi
c 43	60.6	2.6	29392	19	AAV15422	Mouse poly Ig rece
44	60.4	2.6	696	24	ABQ96847	Mouse ES cell rela
c 45	59.8	2.5	347	24	ABQ97071	Mouse ES cell rela

ALIGNMENTS

RESULT 1
ABV73020
ID ABV73020 standard; DNA; 2358 BP.
XX
AC ABV73020;
XX
DT 08-JAN-2003 (first entry)
XX
DE Mouse CARP protein coding sequence upstream DNA fragment.
XX
KW Cardiac ankyrin repeat protein; CARP; cardiant; immunosuppressive;
KW antiinflammatory; gene therapy; antisense gene therapy; mouse; ds.
XX
OS Mus musculus.
XX
PN WO200246220-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001WO-EP15412.
XX
PR 07-DEC-2000; 2000US-251582P.
XX
PA (AVET) AVENTIS PHARMA SA.
PA (REGC) UNIV CALIFORNIA.
PA (BENO/) BENOIT P.
XX
PI Schwartz B, Branellec D, Chien K;
XX
DR WPI; 2002-740642/80.
XX

PT New promoter sequence derived from a portion upstream of the coding
PT sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
PT the level and specificity of expression of a transgene in cardiac
PT muscle cells -
XX
PS Claim 1; Fig 1; 48pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising a fragment of a
CC sequence upstream of the coding part of the gene for the Cardiac Ankyrin
CC Repeat Protein (CARP). (I) is capable of inducing a specific expression
CC in vivo of a gene operably linked to (I), in cardiac cells. (I) or a
CC vector (IIb) comprising (I) is useful for the manufacture of a medicament
CC intended for the treatment of cardiac insufficiency, cardiac hypertrophy
CC and hypoxia, and for preventing rejection during cardiac transplant. An
CC expression cassette under the control of (I) is useful for encoding a
CC protein or RNA which is capable of activating the growth of cardiac
CC cells, reducing or suppressing an immune response, inducing angiogenesis,
CC correcting muscle contractility, cardiac hypertrophy, cardiac
CC insufficiency and myocarditis. (IIb) is useful for expressing a gene of
CC therapeutic interest in vivo, by isolating (IIb) and introducing (IIb) in
CC the cardiac tissue, under conditions so that the gene of interest is
CC expressed. (I), the vectors and the compositions are useful in clinical,
CC experimental, therapeutic and diagnostic fields, and in the treatment and
CC prevention of cardiac pathologies. (I) is also useful for generating
CC transgenic animals which constitute models for studying certain cardiac
CC pathologies. The transgenic animals are also useful for screening
CC molecules for their activity on the regulatory sequences of the gene
CC encoding the CARP protein. The present sequence represents the DNA
CC fragment upstream of the coding sequence of a mouse CARP protein.
XX
SQ Sequence 2358 BP; 636 A; 567 C; 542 G; 613 T; 0 other;

Query Match 100.0%; Score 2358; DB 24; Length 2358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180

Qy        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240

Qy        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300

Qy        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360

Qy        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420

Qy        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480

Qy        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540

Qy        541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600

Qy        601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660

Qy        661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720

Qy        721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780

Qy        781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840

Qy        841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900

Qy        901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960

Qy        961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020

Qy       1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080

Qy       1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140

Qy       1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200

Qy       1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260

Qy       1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320

Qy       1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380

Qy       1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440

Qy       1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500

Qy       1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560

Qy       1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620

Qy       1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680

Qy       1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740

Qy       1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800

Qy       1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860

Qy       1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920

Qy       1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980

Qy       1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040

Qy       2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100

Qy       2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160

Qy       2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220

Qy       2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280

Qy       2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340

Qy       2341 GGTCGGAGGCCACCATGG 2358
             ||||||||||||||||||
Db       2341 GGTCGGAGGCCACCATGG 2358
```

RESULT 2
AAA10406
ID AAA10406 standard; DNA; 2247 BP.
XX
AC AAA10406;
XX
DT 18-JUL-2000 (first entry)
XX
DE Murine cardiac ankyrin repeat protein (CARP) promoter.
XX
KW Cardiac ankyrin repeat protein; CARP promoter; murine; adenovirus vector;
KW cardiac specific; heart disease; gene therapy; ds.
XX
OS Mus musculus.
XX
PN WO200015821-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-US20730.
XX
PR 11-SEP-1998; 98US-0099960.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chien KR, Wang Y, Evans S;
XX
DR WPI; 2000-271457/23.
XX
PT Human type-5 recombinant adenovirus vector used for targeted gene
PT therapy for heart disease and evaluating gene function contains a
PT tissue-restricted promoter and inverted terminal repeat sequences -
XX
PS Claim 8; Page 29-30; 33pp; English.
XX
CC The invention relates to a human type-5 recombinant adenovirus vector
CC for achieving cardiac-restricted transcription of a gene of interest.
CC The vector comprises inverted terminal repeat (ITR) sequences from human
CC adeno-associated virus (AAV) type 2 (AAA10404-A10405) and a cardiac
CC tissue- specific promoter. In particular, the promoter is that of the
CC cardiomyocyte- restricted cardiac ankyrin repeat protein (CARP) gene.
CC The adenovirus vector is used for targetted gene therapy for heart
CC disease and for evaluating gene function. Cardiac restricted
CC transcription of a transgene in both neonatal and mature cardiac tissues
CC can be achieved to treat inherited and acquired heart diseases. The
CC vector is suitable for tissue-specific use in vivo and in vitro and
CC provides cardiac restricted transcription. The present sequence
CC represents the murine cardiac ankyrin repeat protein (CARP) promoter.
XX
SQ Sequence 2247 BP; 609 A; 549 C; 485 G; 589 T; 15 other;

Query Match 74.8%; Score 1763.4; DB 21; Length 2247;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 2124; Conservative 0; Mismatches 69; Indels 137; Gaps 19;

```
Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 90

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 150

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 210

Qy        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 270

Qy        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 330

Qy        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 390

Qy        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 450

Qy        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 510

Qy        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 570

Qy        541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 630
```

```
Qy       601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660
Db       631 CACCTTCTA-CTCTGCATCCATAGCAAGTAGCCTAATG-TTCTGNGTCTAGG-GTCATCT 687

Qy       661 CTGTGAATCGAGATCCTTGG-CCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATT 719
Db       688 CTGTGAATCGAGATCCTTGGCCCTTGTTTGAATTAGGGAGGCACAAAATCTTAAAAAATT 747

Qy       720 CAAGACTGCTCAGCAGCCCAGA--GTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCC 777
Db       748 CAAGACTGNTCAACAANCCANAAGTCCTTTCTCAAAAGGAAAGGNCTTAACTNTNANCCC 807

Qy       778 CCCTTAGCTCT--GAGTCAGGCCTGGAAC--AAACGGCCACAGGAATGAGAAAAGC---- 829
Db       808 CCCTTTACTTTTGAGTCAAGGCCTGGAACCAAACCGGCCCCAGGAATGAAAAAAGCTTGC 867

Qy       830 --TGCCATAGCTGCTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACG 887
Db       868 CATNACCTGGTTGGCCCCTTTNAANAGGNCAAAAAAAAATTGTGGTTAACNTTGAAAAAC 927

Qy       888 AGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTA 947
Db       928 CGAAGACCAACAGTTATCC---------------------------------------- 946

Qy       948 GGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTG 1007
Db       947 ------------------------------------------------------------ 946

Qy      1008 TACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCT 1067
Db       947 ----------TCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCT 996

Qy      1068 TACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACA 1127
Db       997 TACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGA-GGCCAGCATAACA 1055

Qy      1128 AACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCT 1187
Db      1056 AACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCT 1115

Qy      1188 CTTTCATAAAAAATAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAA 1247
Db      1116 CTTTCATAAAAAATAAAAAAAGTATTTTA-GTGGCTCTTACGATAGAATCTTTCCTCGAA 1174

Qy      1248 CTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCC 1307
Db      1175 CTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCC 1234

Qy      1308 AGAAACAAGTATTTTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGT 1367
Db      1235 AGAAACAAG-ATTTTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGT 1293

Qy      1368 TAGGAATAGAACACGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAG 1427
Db      1294 TAGGAATAGAACACGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAG 1353

Qy      1428 TCTCCAGCCTCAGACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGG 1487
Db      1354 TCTCCAGCCTCAGACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGG 1413

Qy      1488 AAACTTTACTGACAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTG 1547
Db      1414 AAACTTTACTGACAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTG 1473

Qy      1548 TGCATCACTTCGGCCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACG 1607
Db      1474 -GCATCACTTCGGCCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACG 1532

Qy      1608 GTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGA 1667
Db      1533 GTGAGCCTGTGGT-CACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGA 1591

Qy      1668 ATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCA 1727
Db      1592 ATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCA 1651

Qy      1728 GACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTT 1787
Db      1652 GACATCA-TCCAGAGAATTCCAAACAGATAG-GACAAGTGGCACCCAGACCCATCTCCTT 1709

Qy      1788 CCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACT 1847
Db      1710 CCCCTCGGGCTGATTATCCCCA-AAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACT 1768

Qy      1848 GGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTT 1907
Db      1769 GGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTT 1828

Qy      1908 GCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATG 1967
Db      1829 GCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATG 1888

Qy      1968 TGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATT 2027
Db      1889 T-GTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATT 1947

Qy      2028 GCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCAT 2087
Db      1948 GCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCAT 2007

Qy      2088 ACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAA 2147
Db      2008 ACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAA 2067

Qy      2148 AATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGG 2207
Db      2068 AATTCACCTGCCTCTGAATTGGCCACTGGTGGGAGCAGGGGTGTGACTTGGCTTCCCAGG 2127

Qy      2208 CTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACA 2267
Db      2128 CTGGAAGATTATCTCACCCAGCCCTA-CTATATAACGGGCTGGTGTGGAGGGGCTCCACA 2186

Qy      2268 GGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGG 2317
Db      2187 GGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAGACTCACGG 2236
```

```
RESULT 3
ABV73021
ID   ABV73021 standard; DNA; 2074 BP.
XX
AC   ABV73021;
XX
DT   08-JAN-2003  (first entry)
XX
DE   Human CARP protein coding sequence upstream DNA fragment.
XX
KW   Cardiac ankyrin repeat protein; CARP; cardiant; immunosuppressive;
KW   antiinflammatory; gene therapy; antisense gene therapy; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246220-A2.
XX
PD   13-JUN-2002.
XX
PF   05-DEC-2001; 2001WO-EP15412.
XX
PR   07-DEC-2000; 2000US-251582P.
XX
PA   (AVET ) AVENTIS PHARMA SA.
PA   (REGC ) UNIV CALIFORNIA.
PA   (BENO/) BENOIT P.
XX
PI   Schwartz B,  Branellec D,  Chien K;
XX
```

```
DR   WPI; 2002-740642/80.
XX
PT   New promoter sequence derived from a portion upstream of the coding
PT   sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
PT   the level and specificity of expression of a transgene in cardiac
PT   muscle cells -
XX
PS   Claim 5; Fig 2; 48pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising a fragment of a
CC   sequence upstream of the coding part of the gene for the Cardiac Ankyrin
CC   Repeat Protein (CARP). (I) is capable of inducing a specific expression
CC   in vivo of a gene operably linked to (I), in cardiac cells. (I) or a
CC   vector (IIb) comprising (I) is useful for the manufacture of a medicament
CC   intended for the treatment of cardiac insufficiency, cardiac hypertrophy
CC   and hypoxia, and for preventing rejection during cardiac transplant. An
CC   expression cassette under the control of (I) is useful for encoding a
CC   protein or RNA which is capable of activating the growth of cardiac
CC   cells, reducing or suppressing an immune response, inducing angiogenesis,
CC   correcting muscle contractility, cardiac hypertrophy, cardiac
CC   insufficiency and myocarditis. (IIb) is useful for expressing a gene of
CC   therapeutic interest in vivo, by isolating (IIb) and introducing (IIb) in
CC   the cardiac tissue, under conditions so that the gene of interest is
CC   expressed. (I), the vectors and the compositions are useful in clinical,
CC   experimental, therapeutic and diagnostic fields, and in the treatment and
CC   prevention of cardiac pathologies. (I) is also useful for generating
CC   transgenic animals which constitute models for studying certain cardiac
CC   pathologies. The transgenic animals are also useful for screening
CC   molecules for their activity on the regulatory sequences of the gene
CC   encoding the CARP protein. The present sequence represents the DNA
CC   fragment upstream of the coding sequence of a human CARP protein.
XX
SQ   Sequence 2074 BP; 612 A; 469 C; 416 G; 572 T; 5 other;

  Query Match             15.6%;  Score 367.8;  DB 24;  Length 2074;
  Best Local Similarity   76.4%;  Pred. No. 5.9e-97;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
             ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
             |||||||||||| ||| |||  ||  |||   |||||||||||| |||| |   |||||
Db       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
             |||||||||||||||||||| |||  ||||| ||||| | |||||||||  |||||
Db       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465

Qy       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
             ||||||  |||||||||||||| |||||||||||||||| |||||||| |     | |
Db       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524

Qy       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837
             |||  ||| | |||     |    ||  |||  | ||||||||||||     |||   | |
Db       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGGACAAGTTTCC 1584

Qy       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897
             |   | || || |||| |||||||||   ||||||||||||||  |||   |||||||
Db       1585 C---CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641

Qy       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955
                 || ||   | |||||||||||| | |||||||||| ||||||||||||  ||||
Db       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698

Qy       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015
             || |||| |||||||||||| | ||||| |||||||||||| | ||  || |  | ||
Db       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757

Qy       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075
             |||||| |||||||||||| |||| ||||  |||||||||| ||||||||||||||||
Db       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817

Qy       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130
             |||||||| ||||  | || |       |   |  ||  |||||||||  ||||||||||
Db       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877

Qy       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190
             ||| ||||||||||||||||||||||||||||||| |||||  ||  ||||||| || |||
Db       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937

Qy       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249
             | |||| | |||||||| ||||||||||||||||||| |||| ||||||||| | | | |
Db       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997

Qy       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309
             |||||||||||  |  |||||||  ||||||| ||| | | || || ||||||||||
Db       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056

Qy       2310 A 2310
              |
Db       2057 A 2057
```

```
RESULT 4
AAI93508
ID   AAI93508 standard; cDNA; 1988 BP.
XX
AC   AAI93508;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13568.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO13577.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 13568; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1988 BP; 632 A; 384 C; 472 G; 500 T; 0 other;

  Query Match          5.5%;  Score 130.2;  DB 22;  Length 1988;
  Best Local Similarity   83.1%;  Pred. No. 3.6e-27;
  Matches  172;  Conservative    0;  Mismatches   33;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  ||||||||||||| | ||||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
               |||||  ||  ||||||| || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | ||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | | || || ||||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 5
AAH02910
ID   AAH02910 standard; DNA; 1901 BP.
XX
AC   AAH02910;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response coding sequence SEQ ID NO: 73.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200125427-A1.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840.
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
PI   Nojima H,  Yoshisue H,  Obayashi M,  Ota T,  Kawabata A,  Sakurada K;
PI   Kuga T,  Sekine S,  Nakamura Y,  Sugano S;
XX
DR   WPI; 2001-266308/27.
DR   P-PSDB; AAB90787.
XX
PT   DNA sequences, proteins encoded by them and antibodies against them
PT   useful in diagnosis and treatment of vascular disease caused by
PT   arteriosclerosis -
XX
PS   Claim 20; Page 422-425; 678pp; Japanese.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human shear stress response proteins. These are useful in the
CC   diagnosis, treatment and screening of vascular diseases caused by
CC   arteriosclerosis, including heart failure, post-PTCA restenosis and
CC   hypertension.
XX
SQ   Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;

  Query Match          5.5%;  Score 128.6;  DB 22;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 1e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  ||||||||||||| | ||||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
               |||||  ||  ||||||  || |||| |||| | |||||||| ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | ||||||||||||| |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | | || || ||||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 6
AAD27217
ID   AAD27217 standard; DNA; 1901 BP.
XX
AC   AAD27217;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human CAA58676 protein encoding EST clone X83703 DNA.
XX
KW   Human; congestive heart failure; dilative cardiomyopathy; sudden death;
KW   hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
KW   heart muscle disease; conduction disorder; coronary heart disease;
KW   systemic arterial hypertension; pulmonary hypertension; endocarditis;
KW   pulmonary heart disease; valvular heart disease; pericardial disease;
KW   congenital heart disease; gene therapy; syncope; transgenic animal;
KW   expressed sequence tag; EST; clone X83703; CAA58676 protein; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             250..1209
FT                   /*tag= a
FT                   /product= "Human CAA58676 protein"
FT   misc_feature    198..431
FT                   /*tag= b
FT                   /note= "66268 cDNA fragment"
FT   misc_feature    1604..1753
FT                   /*tag= b
FT                   /note= "S1MC01-1 cDNA fragment"
FT   misc_feature    1302..1306
FT                   /*tag= b
FT                   /note= "AU-rich mRNA decay element"
FT   misc_feature    1391..1401
FT                   /*tag= b
FT                   /note= "AU-rich mRNA decay element"
FT   misc_feature    1415..1423
FT                   /*tag= b
FT                   /note= "AU-rich mRNA decay element"
XX
PN   WO200192567-A2.
XX
PD   06-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-EP06165.
XX
PR   30-MAY-2000; 2000US-207400P.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bunk D,  Reuner B,  Beck J,  Henkel T;
XX
DR   WPI; 2002-122073/16.
```

```
DR   P-PSDB; AAE16633.
XX
PT   Identifying a subject at risk for a heart disease e.g. congestive heart
PT   failure, dilative cardiomyopathy, heart muscle disease, by quantifying
PT   the polypeptide expressed by genes abnormally expressed in heart tissue
PT   -
XX
PS   Claim 2a; Fig 10b; 154pp; English.
XX
CC   The patent discloses novel target genes abnormally expressed in heart
CC   tissues and their corresponding proteins. The invention also relates to
CC   methods for assessing the expression level of these genes. The method
CC   is used for testing the predisposition of mammals and preferably humans
CC   for a heart disease or for an acute state of such a disease. It is also
CC   useful to treat diseases of the heart such as congestive heart failure,
CC   dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
CC   myopathy, specific heart muscle disease, rhythm and conduction disorders,
CC   syncope and sudden death, coronary heart disease, systemic arterial
CC   hypertension, pulmonary hypertension, pulmonary heart disease, valvular
CC   heart disease, congenital heart disease, pericardial disease and
CC   endocarditis. Sequences of the invention are also used in gene therapy.
CC   A transgenic non-human mammal comprising the sequences of the invention
CC   are useful for the development for medicaments for the treatments of
CC   heart diseases. The present DNA sequence is expressed sequence tag
CC   (EST) clone X83703 which encodes CAA58676 protein.
XX
SQ   Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;

  Query Match             5.5%;  Score 128.6;  DB 24;  Length 1901;
  Best Local Similarity   82.6%;  Pred. No. 1e-26;
  Matches  171;  Conservative    0;  Mismatches   34;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
             ||  |||||||||  ||||||||||||||| | ||||||||||||||||||||||||||||
Db         27 TCCCCTCCCTCTTCAGCTTCCCAGACACTGATTCTGGAATGAAAATTCACCTGCCTCTGA 86

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
              |||||  ||  ||||||  || |||| |||| | ||||||||  ||||||||||||||||
Db         87 GTTGGCTCCTAATGGGGGTGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 146

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
             || |||| |||||||||| | | | ||||||||||||||  |  ||||||||  |||||||
Db        147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 206

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
             ||| | | || || ||||||||||| |
Db        207 TTCCTTC-CACGACAGAAAAACATACA 232


RESULT 7
ABZ24607
ID   ABZ24607 standard; DNA; 7280 BP.
XX
AC   ABZ24607;
XX
DT   07-APR-2003  (first entry)
XX
DE   Mouse gastric antrum mucosal protein-18 gene.
XX
KW   Mouse; gastrokine; AMP-18; gastric antrum mucosal protein; mitogen;
KW   growth factor; vulnerary; antiulcer; gene; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..1874
FT                   /*tag=  a
FT   CDS             1945..6542
FT                   /*tag=  b
FT                   /product= "Mouse pre-AMP-18"
FT                   /note= "contains introns"
FT   exon            1906..1956
FT                   /*tag=  c
FT                   /number= 1
FT   intron          1957..3531
FT                   /*tag=  d
FT                   /number= 1
FT   exon            3532..3582
FT                   /*tag=  e
FT                   /number= 2
FT   intron          3583..3672
FT                   /*tag=  f
FT                   /number= 2
FT   exon            3673..3813
FT                   /*tag=  g
FT                   /number= 3
FT   intron          3814..4594
FT                   /*tag=  h
FT                   /number= 3
FT   exon            4595..4705
FT                   /*tag=  i
FT                   /number= 4
FT   intron          4706..5607
FT                   /*tag=  j
FT                   /number= 4
FT   exon            5608..5749
FT                   /*tag=  k
FT                   /number= 5
FT   intron          5750..6444
FT                   /*tag=  l
FT                   /number= 5
FT   exon            6445..6542
FT                   /*tag=  m
FT                   /number= 6
FT   polyA_signal    6636..6641
FT                   /*tag=' n
XX
PN   WO200292758-A2.
XX
PD   21-NOV-2002.
XX
PF   29-MAR-2002; 2002WO-US10148.
XX
PR   29-MAR-2001; 2001US-0821726.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Toback FG,  Martin TE,  Powell CT,  Agarwal K;
XX
DR   WPI; 2003-120666/11.
DR   P-PSDB; ABP58258.
XX
PT   Gastric Antrum Mucosal Protein 18, useful for preparing a composition
PT   for healing of the injured gastrointestinal tract  -
XX
PS   Disclosure; Fig 4; 67pp; English.
XX
CC   The present sequence is that of a gene encoding a novel member of
CC   the murine gastrokine group, designated antrum mucosal protein 18
CC   (AMP-18). AMP-18 was initially detected in mammalian gastric
CC   antrum mucosa by a differential screen of cDNA libraries obtained
CC   from different regions of the pig stomach. A cDNA was also
CC   isolated from a mouse library (see ABZ24608). Genomic AMP-18 DNA
CC   sequences were subsequently cloned as a prelude to the analysis of
CC   gene regulatory elements. AMP-18 protein is expressed at high
CC   levels only in the gastric antrum, and is synthesised in the
CC   lumenal surface mucosal cells. Partially purified AMP-18s from
CC   mouse and pig antrum tissue are mitogenic to confluent stomach
CC   and kidney epithelial cells in culture. This effect is inhibited
CC   by specific antisera, suggesting that AMP-18, or its products, is a
CC   growth factor. AMP-18 may play a critical role in the repair of
CC   the stomach epithelium following damage by agents such as alcohol,
CC   non-steroidal antiinflammatory drugs, or pathogens, particularly
CC   Helicobacter pylori, a causative agent of gastric ulcers and
CC   possibly cancers. AMP-18 protein, or an active peptide of AMP-18,
CC   is used in a claimed method of stimulating the growth of epithelial
```

```
CC   cells in the gastrointestinal tract.
XX
SQ   Sequence 7280 BP; 2131 A; 1597 C; 1606 G; 1944 T; 2 other;

  Query Match             4.0%;  Score 95;  DB 25;  Length 7280;
  Best Local Similarity   80.5%;  Pred. No. 1.6e-16;
  Matches  136;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    2;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
              ||||| ||||  ||||||| |||| |||||  || ||||||||||  |  |||||||||
Db        391 TCTAGAGCAGTGGTTCTCATCCTGTGGGCCATGAGCCCTTTGGGGGGGTTGAACGACCCT 450

Qy        258 TT-ACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAAC 316
              || ||||||||||||||||  ||||||  || | || ||||||||||| ||||| ||||
Db        451 TTCACAGGGGTCACATATCAGATATCCTGCATCTTAGCTATTTACATTATGATTCATAAC 510

Qy        317 AGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTG 365
              ||||||||||||  || || ||| ||| ||  ||||||  ||||| |||
Db        511 AGTAGCAAAATT--AGTTAGGAAGTAGGAACAAAATAACGTTATGGTTG 557


RESULT 8
AAD50815
ID   AAD50815 standard; DNA; 7280 BP.
XX
AC   AAD50815;
XX
DT   02-APR-2003  (first entry)
XX
DE   Mouse pre-AMP-18 gene.
XX
KW   Cellular growth stimulating protein; gastric antrum mucosal protein;
KW   gastrokine; AMP-18 protein; gastro-intestinal disorder; cell therapy;
KW   ulcer; mouse; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   exon            1906..1956
FT                   /*tag=  a
FT                   /number= 1
FT   intron          1957..3531
FT                   /*tag=  b
FT                   /number= 1
FT   exon            3532..3582
FT                   /*tag=  c
FT                   /number= 2
FT   intron          3583..3672
FT                   /*tag=  d
FT                   /number= 2
FT   exon            3673..3813
FT                   /*tag=  e
FT                   /number= 3
FT   intron          3814..4594
FT                   /*tag=  f
FT                   /number= 3
FT   exon            4595..4705
FT                   /*tag=  g
FT                   /number= 4
FT   intron          4706..5607
FT                   /*tag=  h
FT                   /number= 4
FT   exon            5608..5749
FT                   /*tag=  i
FT                   /number= 5
FT   intron          5750..6444
FT                   /*tag=  j
FT                   /number= 5
FT   exon            6445..6542
FT                   /*tag=  k
FT                   /number= 6
FT   polyA_site      6636
FT                   /*tag=  l
XX
PN   WO200278640-A2.
XX
PD   10-OCT-2002.
XX
PF   29-MAR-2002; 2002WO-US09885.
XX
PR   29-MAR-2001; 2001US-0821726.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Toback GF,  Martin TE,  Walsh-Reitz M;
XX
DR   WPI; 2003-103239/09.
XX
PT   Protein inhibitor, useful for treating gastrointestinal disorders or
PT   diseases comprises isolated homologous cellular stimulating proteins -
XX
PS   Disclosure; Fig 4; 84pp; English.
XX
CC   The invention relates to a protein inhibitor which comprises homologous
CC   cellular growth stimulating proteins designated gastrokines. The
CC   invention also provides gastric antrum mucosal proteins designated
CC   AMP-18 which belongs to the novel group of gastrokines and nucleic acid
CC   molecules encoding such proteins. Pharmaceutical composition comprising
CC   growth stimulating peptide derived from a gastrokine protein is useful
CC   for treating gastro-intestinal disorder or diseases associated with
CC   overgrowth of gastric epithelia e.g. ulcer. The invention is useful in
CC   cell therapy. The present sequence is mouse pre-AMP-18 gene.
XX
SQ   Sequence 7280 BP; 2131 A; 1597 C; 1606 G; 1944 T; 2 other;

  Query Match             4.0%;  Score 95;  DB 25;  Length 7280;
  Best Local Similarity   80.5%;  Pred. No. 1.6e-16;
  Matches  136;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    2;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
              ||||| ||||  ||||||| |||| |||||  || ||||||||||  |  |||||||||
Db        391 TCTAGAGCAGTGGTTCTCATCCTGTGGGCCATGAGCCCTTTGGGGGGGTTGAACGACCCT 450

Qy        258 TT-ACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAAC 316
              || ||||||||||||||||  ||||||  || | || ||||||||||| ||||| ||||
Db        451 TTCACAGGGGTCACATATCAGATATCCTGCATCTTAGCTATTTACATTATGATTCATAAC 510

Qy        317 AGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTG 365
              ||||||||||||  || || ||| ||| ||  ||||||  ||||| |||
Db        511 AGTAGCAAAATT--AGTTAGGAAGTAGGAACAAAATAACGTTATGGTTG 557


RESULT 9
ABZ82337/c
ID   ABZ82337 standard; DNA; 18105 BP.
XX
AC   ABZ82337;
XX
DT   15-MAY-2003  (first entry)
XX
DE   Mouse histamine receptor H3 gene.
XX
KW   Mouse; drug composition; food intake disorder; body weight disorder;
KW   histamine receptor H3; anorectic; gene; ds.
XX
OS   Mus musculus.
XX
PN   WO2003004637-A1.
XX
PD   16-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-JP06580.
XX
PR   02-JUL-2001; 2001JP-0201413.
XX
PA   (BANY ) BANYU PHARM CO LTD.
```

```
XX
PI   Kotani H,  Takahashi K,  Suwa H;
XX
DR   WPI; 2003-221596/21.
XX
PT   Histamine receptor H3 protein and its agonists and antagonists for
PT   treatment and prevention of body weight and food intake disorders -
XX
PS   Example 1; Page 53-65; 73pp; Japanese.
XX
CC   The invention relates to novel drug compositions for the treatment and
CC   prevention of disorders of food intake and body weight containing as the
CC   active component histamine receptor H3 protein, or DNA encoding it, or an
CC   agonist or antagonist to it. A composition of the invention has anorectic
CC   activity. Compositions containing the histamine receptor H3 protein may
CC   be used in the treatment and prevention of disorders of body weight and
CC   food intake. The present sequence represents the mouse histamine receptor
CC   H3 gene.
XX
SQ   Sequence 18105 BP; 3981 A; 5123 C; 4944 G; 4041 T; 16 other;

  Query Match             3.8%;  Score 89.2;  DB 25;  Length 18105;
  Best Local Similarity   79.1%;  Pred. No. 1.3e-14;
  Matches  106;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
              | |||| ||   |||||||||||||||  ||||||   | |||| | |||| | |  ||||
Db      10825 CAACCCCTTTGGGGGGTCACATATCAGATATCCTGCCTATCAGATGTTTATACTGTGATT 10766

Qy        311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
              ||| ||||||||||||||||  |||||| ||||||  ||||||||||||| |||  |||
Db      10765 CGTTACAGTAGCAAAATTACACTTATGAAGTAGCAACAAAATAATTTTATGGTTGGGGGT 10706

Qy        371 CACCACAACATGAG 384
              |||||||||| ||
Db      10705 AACCACAACATAAG 10692


RESULT 10
ABX16390
ID   ABX16390 standard; DNA; 659158 BP.
XX
AC   ABX16390;
XX
DT   09-APR-2003  (first entry)
XX
DE   Mouse high growth region.
XX
KW   High growth region; high growth phenotype; Socs2; body size;
KW   supressor of cytokine signaling 2; ds; mouse.
XX
OS   Mus sp.
XX
PN   US2002155564-A1.
XX
PD   24-OCT-2002.
XX
PF   26-JAN-2001; 2001US-0771208.
XX
PR   29-DEC-1997;   97US-0999477.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Medrano JF,  Bradford E,  Horvat S;
XX
DR   WPI; 2003-182637/18.
XX
PT   Novel gene that when downregulated or knocked-out, results in high
PT   growth phenotype, useful for regulating body size in mammals e.g.
PT   rodent, bovine and canine -
XX
PS   Disclosure; SEQ ID NO 18; 49pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule encoding a
CC   gene product that, when knocked out, results in a high growth (hg)
CC   phenotype. For example a nucleic acid disrupting the Socs2 gene is
CC   useful for producing an animal characterised by a hg phenotype, by
CC   inhibiting expression of Socs2 (supressor of cytokine siganling 2)
CC   gene. The nucleic acids of the invention are useful for regulating
CC   body size in mammals. gene. The nucleic acids of the invention
CC   are useful for regulating body size in mammals. This sequence
CC   represents the mouse high growth region.
CC   Note: This sequence did not form part of the printed specification
CC   but was obtained in electronic format directly from the US patent
CC   office at
CC   seqdata.uspto.gov/sequence.html?DocID=20020155564.
XX
SQ   Sequence 659158 BP; 177521 A; 147222 C; 149414 G; 184754 T; 247 other;

  Query Match             3.8%;  Score 88.8;  DB 25;  Length 659158;
  Best Local Similarity   74.7%;  Pred. No. 1.3e-13;
  Matches  139;  Conservative    0;  Mismatches   42;  Indels    5;  Gaps    2;

Qy        204 GCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAG 263
              |||||||||| |||| || ||| |  || | ||||||| |    ||| | || || ||||
Db     208828 GCAGCAGTTCCCAACGTGTGGGTCCAGATCTCTTTGGGAG-CCGAAAGGGCCTTTCACAG 208886

Qy        264 GGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCA 323
              ||||||||  |||  || |||  ||||    |||||| ||||| |||| || |||  |||
Db     208887 GGGTCACAGGTCAGATACCCTGAATGTTGCATATTTATATTACAATTCATAGCAGCTGCA 208946

Qy        324 AAATTACAGGTATGAAATAGCAATGAAATAATTTTA----TGATTGAAGGTCACCACAAC 379
              ||| ||||| |||||| |||||||||||||||| |||    ||   | ||||||||||| |
Db     208947 AAACTACAGCTATGAAGTAGCAATGAAATAATCTTACAGTTGGGGGGAGGTCACCACAGC 209006

Qy        380 ATGAGG 385
              ||||||
Db     209007 ATGAGG 209012


RESULT 11
AAX55300/c
ID   AAX55300 standard; DNA; 48974 BP.
XX
AC   AAX55300;
XX
DT   19-JUL-1999  (first entry)
XX
DE   Mouse Presenilin-1 gene sequence.
XX
KW   Neuron-specific transcription; promoter; mouse; Presenilin-1 gene;
KW   transgenic; laboratory model; Alzheimer's disease; ss.
XX
OS   Mus sp.
XX
PN   WO9910368-A1.
XX
PD   04-MAR-1999.
XX
PF   28-AUG-1998;   98WO-US17905.
XX
PR   29-AUG-1997;   97US-0920422.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Mitsuda N,  Roses AD,  Vitek PM;
XX
DR   WPI; 1999-312401/26.
XX
PT   Presenilin-1 gene promoter
XX
PS   Claim 1; Page 51-81; 85pp; English.
XX
CC   The invention provides a DNA molecule that directs neuron-specific
CC   transcription of a promoter segment in a mammalian cell. The promoter
CC   segment is from the mouse genomic Presenilin-1 gene. Transgenic non-human
```

CC mammals containing a DNA expression cassette comprising the neuron-
CC specific promoter are useful as laboratory models for studying the
CC function of the Presenilin gene, and for studying the etiology of
CC Alzheimer's disease.
XX
SQ Sequence 48974 BP; 12579 A; 10749 C; 11189 G; 14438 T; 19 other;

Query Match 3.8%; Score 88.6; DB 20; Length 48974;
Best Local Similarity 69.2%; Pred. No. 3.5e-14;
Matches 135; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

Qy 191 GCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAA 250
Db 15385 GCTACACTCTAGATCGGTGGTTCTCAAGCTGGGGGTCGCGACCCCTTTGGGGGTTGGACA 15326

Qy 251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
Db 15325 A-CCTTTTCACAAGGGTCACACATCAGAGATCCTGCATATCCGATATTTATGTTATGATT 15267

Qy 311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
Db 15266 CATAACAGCAACAGAATTACAGTTAGGAAGTAGAAATGAAAATAATGTTATGATTGGGGC 15207

Qy 371 CACCACAACATGAGG 385
Db 15206 CACCATGACATGAAG 15192

RESULT 12
ABT09109
ID ABT09109 standard; DNA; 707 BP.
XX
AC ABT09109;
XX
DT 05-DEC-2002 (first entry)
XX
DE Phase-1 Rat CT gene SEQ ID No 197.
XX
KW Rat; toxicity study; rat toxic response gene; toxicological response;
KW drug development; phase-1 rat CT gene; ds.
XX
OS Rattus sp.
XX
PN WO200266682-A2.
XX
PD 29-AUG-2002.
XX
PF 29-JAN-2002; 2002WO-US02935.
XX
PR 29-JAN-2001; 2001US-264933P.
PR 26-JUL-2001; 2001US-308161P.
XX
PA (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX
PI Farris G, Hicken SH, Farr SB;
XX
DR WPI; 2002-674961/72.
XX
PT Evaluating the toxicity of an agent, useful in drug development or in
PT determining toxicological responses to a new drug, by determining the
PT expression of rat toxicologically relevant genes in the test animal in
PT response to the test agent -
XX
PS Disclosure; Page 155; 388pp; English.
XX
CC The invention relates to a method used for evaluating the toxicity of an
CC agent comprising determining the expression of a rat toxic response
CC gene(s) in the test animal in response to the agent. The method is useful
CC in drug development, particularly for conducting toxicity studies and
CC analysis before a new drug or compound is approved for human consumption
CC or use. The method is also useful in determining toxicological responses
CC to a new drug. This polynucleotide sequence represents a phase-1 rat CT
CC gene of the invention.
XX
SQ Sequence 707 BP; 205 A; 174 C; 156 G; 169 T; 3 other;

Query Match 3.5%; Score 83.6; DB 24; Length 707;
Best Local Similarity 67.3%; Pred. No. 9.9e-14;
Matches 148; Conservative 0; Mismatches 69; Indels 3; Gaps 2;

Qy 171 GAGCTTCTCTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCG 230
Db 232 GAGCCATCCCTCCAGCCCCAGCCTGTTTTTATGGAAGTGATTCTCAACTCATGGGTCATG 291

Qy 231 ACCCCTTTGGGGG--AATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATAT 288
Db 292 ACCCCTTTGGGGGTTAAATGACCCTTTCACATATCAAATATCAAATCAAATACCCTGCAG 351

Qy 289 GTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATG 348
Db 352 AGCAGATATTCACATTGCAATCCGCAACAGCAGCAAAATTACAGTTACGAAGTAGCAAAG 411

Qy 349 -AAATAATTTTATGATTGAAGGTCACCACAACATGAGGCC 387
Db 412 AAAATAATCTTACGGTTGAAGGTCATCACAACACGAGGAC 451

RESULT 13
ABX16390/c
ID ABX16390 standard; DNA; 659158 BP.
XX
AC ABX16390;
XX
DT 09-APR-2003 (first entry)
XX
DE Mouse high growth region.
XX
KW High growth region; high growth phenotype; Socs2; body size;
KW supressor of cytokine signaling 2; ds; mouse.
XX
OS Mus sp.
XX
PN US2002155564-A1.
XX
PD 24-OCT-2002.
XX
PF 26-JAN-2001; 2001US-0771208.
XX
PR 29-DEC-1997; 97US-0999477.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Medrano JF, Bradford E, Horvat S;
XX
DR WPI; 2003-182637/18.
XX
PT Novel gene that when downregulated or knocked-out, results in high
PT growth phenotype, useful for regulating body size in mammals e.g.
PT rodent, bovine and canine -
XX
PS Disclosure; SEQ ID NO 18; 49pp; English.
XX
CC The invention describes an isolated nucleic acid molecule encoding a
CC gene product that, when knocked out, results in a high growth (hg)
CC phenotype. For example a nucleic acid disrupting the Socs2 gene is
CC useful for producing an animal characterised by a hg phenotype, by
CC inhibiting expression of Socs2 (supressor of cytokine siganling 2)
CC gene. The nucleic acids of the invention are useful for regulating
CC body size in mammals. gene. The nucleic acids of the invention
CC are useful for regulating body size in mammals. This sequence
CC represents the mouse high growth region.
CC Note: This sequence did not form part of the printed specification
CC but was obtained in electronic format directly from the US patent
CC office at
CC seqdata.uspto.gov/sequence.html?DocID=20020155564.
XX

```
SQ   Sequence 659158 BP; 177521 A; 147222 C; 149414 G; 184754 T; 247 other;

  Query Match             3.5%;  Score 82;  DB 25;  Length 659158;
  Best Local Similarity   71.8%;  Pred. No. 1.3e-11;
  Matches  135;  Conservative    0;  Mismatches   50;  Indels    3;  Gaps    2;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
              |||||   ||  ||||||||||  |||  || ||| |||||||  |  |||  |   |
Db     439418 TCTAGACTAGTGGTTCTCAACCTATGGGTTTCAACCTCTTTGGGAGTTTCATATCAGATA 439359

Qy        258 TTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACA 317
              ||    | ||  |||||||   |||||  || |||| ||||| ||||| ||  | |||
Db     439358 TT--CTGCGTAGCATATCAGACATCCTGCATATCAGATATTTCCATTATGACACATAATG 439301

Qy        318 GTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACA 377
              |||||||||| |||| |||||  || |||||||| |||||||| |||  |||||||  ||
Db     439300 GTAGCAAAATCACAGTTATGAGGTAACAATGAAATGATTTTATGGTTG-GGGTCACCGCA 439242

Qy        378 ACATGAGG 385
              |||||  ||
Db     439241 ACATGTGG 439234


RESULT 14
ABA97708
ID   ABA97708 standard; DNA; 10917 BP.
XX
AC   ABA97708;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Wild-type mouse Wolfram Syndrome 1 gene.
XX
KW   Wolfram Syndrome 1; WFS1; transgenic; wolframin; brain; depression;
KW   stress-induced neurochemical change; behavioural change; mouse;
KW   drug screening; antidepressant; gene; ds.
XX
OS   Mus musculus.
XX
PN   WO200191548-A2.
XX
PD   06-DEC-2001.
XX
PF   31-MAY-2001; 2001WO-US17652.
XX
PR   01-JUN-2000; 2000US-209394P.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Roberds SL,  Huff RM;
XX
DR   WPI; 2002-089969/12.
XX
PT   New transgenic non-human animals (mice), useful as models for
PT   depression, for monitoring the efficacy of a drug against depression,
PT   and for screening antidepressants, drugs or genes for ameliorating or
PT   treating depression -
XX
PS   Claim 19; Page 60-63; 65pp; English.
XX
CC   The present sequence represents the genomic DNA sequence of the wild-type
CC   mouse Wolfram Syndrome 1 (WFS1) gene used in the method of the invention.
CC   The specification describes a non-human transgenic mammal comprising a
CC   genome containing a single copy of a wild-type WFS1 gene and a modified
CC   WFS1 allele containing a mutation that disrupts the function of
CC   wolframin, or its transgenic progeny. Wolframin is a predicted
CC   transmembrane protein expressed in many tissues including pancreas and
CC   brain. The transgenic nonhuman animals are useful as models for
CC   depression, particularly as models for depression in humans. The animals
CC   exhibit chronic stress-induced neurochemical and behavioural changes
CC   associated with depression. The transgenic animals are also useful for
CC   screening or identifying antidepressant agents, drugs or genes that may
CC   be employed to ameliorate or treat depression. The animals are also
CC   useful for monitoring the efficacy of a drug against depression.
XX
SQ   Sequence 10917 BP; 2423 A; 2842 C; 2929 G; 2706 T; 17 other;

  Query Match             3.4%;  Score 80.4;  DB 24;  Length 10917;
  Best Local Similarity   70.8%;  Pred. No. 3.9e-12;
  Matches  136;  Conservative    0;  Mismatches   51;  Indels    5;  Gaps    2;

Qy        197 TTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCC 256
              | | ||| |||  |||||||||||| ||| | | ||||||||||||    |  || |||
Db     10340 TGCGAGGACAGAGGTTCTCAACCTGTGGGGCACAACCCCTTTGGGG--TTTGAATGACTT 10397

Qy        257 TTTACAGGGGTCACATATCA---TCTATCCTATATGTCAGGTATTTACATTACGATTCGT 313
              || ||||||||  | |  ||    | |||  | |   ||||||||||||||| | |||| |
Db     10398 TTCACAGGGGTTGCTTGCCAAAGACCATCAGAAAACACAGGTATTTACATTCCAATTCAT 10457

Qy        314 AACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCAC 373
               ||||||| |||||||||| |||| | ||||||||||  | |||     |   ||| ||
Db     10458 GACAGTAGTAAAATTACAGTTATGCAGTAGCAATGAAAATAATTTTATGGTTGGGGTGAC 10517

Qy        374 CACAACATGAGG 385
              |||| ||||| |
Db     10518 CACATCATGAAG 10529


RESULT 15
AAV06155
ID   AAV06155 standard; DNA; 892 BP.
XX
AC   AAV06155;
XX
DT   06-MAY-1998  (first entry)
XX
DE   Viral infection gene SEQ ID NO:75.
XX
KW   Viral infection; tumour suppressor; cellular gene; rat; cancer;
KW   serum protein; inhibitor; malignant phenotype; HIV; influenza;
KW   hepatitis; retrovirus; immunodeficiency; ds.
XX
OS   Rattus sp.
XX
PN   WO9739119-A1.
XX
PD   23-OCT-1997.
XX
PF   11-APR-1997;   97WO-US06067.
XX
PR   15-APR-1996;   96US-0015334.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Dubois RN,  Organ EL,  Rubin DH;
XX
DR   WPI; 1997-526456/48.
XX
PT   Genes involved in viral infection and tumour suppression - used to
PT   develop products for reducing or preventing viral infection or for
PT   suppressing tumours
XX
PS   Claim 1; Page 82-83; 101pp; English.
XX
CC   The present sequence represents a viral infection gene. The present
CC   invention describes nucleic acid sequences isolated from rat. The
CC   sequences of the invention comprise 70 viral infection (VI) genes and
CC   8 tumour suppressor (TS) genes. Propagating cell cultures in the
CC   absence of the serum protein (SP) allows selective elimination of cells
CC   persistently infected with a virus from the cell culture. Inhibitors of
CC   the SP can be used for reducing or inhibiting a viral infection.
CC   Inhibitors of the TS gene products can be used to suppress a malignant
CC   phenotype (MP). The methods and inhibitors can be used with viruses
CC   such as HIV, influenza, hepatitis virus or animal retroviruses such as
CC   simian immunodeficiency virus, avian immunodeficiency virus, bovine
CC   immunodeficiency virus, feline immunodeficiency virus, equine infectious
```

```
CC   anemia virus, caprine arthritis encephalitis virus or visna virus.
CC   Because the identified genes are non-essential to cell survival, the
CC   treatment methods can be used in subjects without serious detrimental
CC   effects to the subjects.
XX
SQ   Sequence 892 BP; 224 A; 206 C; 185 G; 229 T; 48 other;

  Query Match             3.3%;  Score 78;  DB 18;  Length 892;
  Best Local Similarity   70.5%;  Pred. No. 5e-12;
  Matches  148;  Conservative    0;  Mismatches   55;  Indels    7;  Gaps    3;

Qy        177 CTCTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCT 236
              | | | ||     || |  | ||||| |||    ||||  ||||| | |   | ||||||
Db        119 CACGTACATAGATAGTCAAAATCTAGAGCACTGTTTCTATACCTGTGAGTTGCAACCCCT 178

Qy        237 TTGGG---GGAATCAAACGACCCT-TTACAGGGGTCACATATCATCTATCCTATATGTCA 292
              |||||    |   ||||| |||||| | |||||||||| || || |  ||||||  || |||
Db        179 TTGGGAGTGCGGTCAAATGACCCTATCACAGGGGTCTCAAATGAGATATCCTGCATATCA 238

Qy        293 GGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAAT 352
                ||||||||||| ||||| ||  |||| || |||||||| |||||| |  ||   ||||
Db        239 AATATTTACATTATGATTCATAGTAGTACCAGAATTACAGTTATGAAGTTACA---AAAT 295

Qy        353 AATTTTATGATTGAAGGTCACCACAACATG 382
              ||||||||   |||  ||||||||||||||
Db        296 AATTTTATAGCTGAGAGTCACCACAACATG 325


Search completed: November 17, 2003, 23:52:49
Job time : 628.73 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:50 ; Search time 137.266 Seconds
(without alignments)
7582.209 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggatcctttcatgtttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	88.6	3.8	48974	3	US-08-920-422-17	Sequence 17, Appl
	2	78	3.3	892	4	US-09-171-209-75	Sequence 75, Appl
	3	74.8	3.2	37950	3	US-09-338-907-183	Sequence 183, App
	4	74.8	3.2	37950	4	US-09-218-207-183	Sequence 183, App
c	5	74.6	3.2	10614	1	US-08-135-511-35	Sequence 35, Appl
c	6	74.6	3.2	10614	1	US-08-187-453-35	Sequence 35, Appl
	7	73.4	3.1	4164	4	US-08-882-164D-38	Sequence 38, Appl
	8	68.6	2.9	4072	3	US-09-272-496-7	Sequence 7, Appli
c	9	68	2.9	16442	3	US-08-781-891-208	Sequence 208, App
c	10	68	2.9	16442	4	US-09-618-166-208	Sequence 208, App
	11	65.6	2.8	29604	3	US-08-781-891-207	Sequence 207, App
	12	65.6	2.8	29604	4	US-09-618-166-207	Sequence 207, App
c	13	65	2.8	6645	2	US-08-380-403A-4	Sequence 4, Appli
c	14	65	2.8	6645	2	US-08-895-628-4	Sequence 4, Appli
c	15	65	2.8	6645	4	US-08-895-810D-4	Sequence 4, Appli
	16	63	2.7	3450	2	US-08-378-617A-9	Sequence 9, Appli
c	17	62	2.6	7208	3	US-09-166-186-107	Sequence 107, App
c	18	62	2.6	7208	3	US-09-313-932-107	Sequence 107, App
c	19	54	2.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	20	50.6	2.1	5764	4	US-09-312-762A-8	Sequence 8, Appli
	21	47	2.0	26700	1	US-08-472-217-1	Sequence 1, Appli
	22	47	2.0	26700	2	US-08-488-199-5	Sequence 5, Appli
	23	47	2.0	26700	3	US-08-760-534A-1	Sequence 1, Appli
	24	47	2.0	26700	4	US-09-336-757-1	Sequence 1, Appli
	25	38.8	1.6	818	4	US-09-328-475C-157	Sequence 157, App
	26	38.6	1.6	2830	1	US-08-604-333-3	Sequence 3, Appli
	27	38.6	1.6	2830	3	US-09-110-618-3	Sequence 3, Appli
	28	38.6	1.6	2830	4	US-09-578-178-3	Sequence 3, Appli
	29	38.6	1.6	2830	4	US-09-577-806-3	Sequence 3, Appli
	30	37.2	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	31	36.8	1.6	277	3	US-09-007-005-3	Sequence 3, Appli
	32	36.8	1.6	277	3	US-09-244-796-3	Sequence 3, Appli
	33	36.6	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	34	36.2	1.5	1209	4	US-08-462-509B-1	Sequence 1, Appli
	35	36.2	1.5	1209	5	PCT-US95-05616-1	Sequence 1, Appli
	36	36.2	1.5	1564	2	US-08-846-705-4	Sequence 4, Appli
	37	36.2	1.5	1564	3	US-08-846-704-1	Sequence 1, Appli
	38	36.2	1.5	1564	3	US-08-846-704-3	Sequence 3, Appli
	39	35	1.5	9573	4	US-09-220-132-168	Sequence 168, App
c	40	35	1.5	80246	3	US-09-078-294-4	Sequence 4, Appli
c	41	35	1.5	80595	3	US-09-078-294-3	Sequence 3, Appli
	42	34.8	1.5	2088	2	US-08-602-264A-1	Sequence 1, Appli
	43	34.8	1.5	2088	3	US-08-461-018A-1	Sequence 1, Appli
	44	34.8	1.5	2088	3	US-09-216-958-1	Sequence 1, Appli
c	45	34.4	1.5	1843	4	US-09-622-540A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-920-422-17/c
; Sequence 17, Application US/08920422A
; Patent No. 6255473
; GENERAL INFORMATION:
; APPLICANT: Vitek, Michael P.
; APPLICANT: Mitsuda, No. 6255473iaki
; APPLICANT: Roses, Allen D.
; TITLE OF INVENTION: Presenilin-1 Gene Promoter
; FILE REFERENCE: VITEKPRESENILIN
; CURRENT APPLICATION NUMBER: US/08/920,422A
; CURRENT FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 48974
; TYPE: DNA
; ORGANISM: Mus musculus
US-08-920-422-17

Query Match 3.8%; Score 88.6; DB 3; Length 48974;
Best Local Similarity 69.2%; Pred. No. 3.9e-16;
Matches 135; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

```
Qy      191 GCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAA 250
            ||   | |||||  | |  |||||||| |||||||| | ||||||||||||||||    | |
Db    15385 GCTACACTCTAGATCGGTGGTTCTCAAGCTGGGGGTCGCGACCCCTTTGGGGGTTGGACA 15326

Qy      251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
              | || ||| |||||||| ||||   |||||  || || | ||||||  ||| ||||
Db    15325 A-CCTTTTCACAAGGGTCACACATCAGAGATCCTGCATATCCGATATTTATGTTATGATT 15267

Qy      311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
            | |||||| | || |||||||| || ||| ||| |||||||  | | |      |   ||
Db    15266 CATAACAGCAACAGAATTACAGTTAGGAAGTAGAAATGAAAATAATGTTATGATTGGGGC 15207

Qy      371 CACCACAACATGAGG 385
            |||||  |||||| |
Db    15206 CACCATGACATGAAG 15192
```

RESULT 2
US-09-171-209-75
; Sequence 75, Application US/09171209
; Patent No. 6448000
; GENERAL INFORMATION:
; APPLICANT: VANDERBILT UNIVERSITY
; 305 Kirkland Hall
; Nashville, TN 37240
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN VIRAL

```
;                          INFECTION
;       NUMBER OF SEQUENCES: 83
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Needle & Rosenberg, P.C.
;            STREET: 127 Peachtree Street, Suite 1200
;            CITY: Atlanta
;            STATE: Georgia
;            COUNTRY: USA
;            ZIP: 30303-1811
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/171,209
;            FILING DATE: 08-Mar-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/US97/06067
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Selby, Elizabeth
;            REGISTRATION NUMBER: 38,298
;            REFERENCE/DOCKET NUMBER: 22000.0061/P
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 404 688 0770
;            TELEFAX: 404 688 9880
;   INFORMATION FOR SEQ ID NO: 75:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 892 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-171-209-75

  Query Match            3.3%;  Score 78;  DB 4;  Length 892;
  Best Local Similarity   70.5%;  Pred. No. 6.9e-14;
  Matches  148;  Conservative    0;  Mismatches   55;  Indels    7;  Gaps    3;

Qy        177 CTCTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCT 236
              | | | ||     || |  | ||||| |||    ||||  ||||| | |    | ||||||
Db        119 CACGTACATAGATAGTCAAAATCTAGAGCACTGTTTCTATACCTGTGAGTTGCAACCCCT 178

Qy        237 TTGGG---GGAATCAAACGACCCT-TTACAGGGGTCACATATCATCTATCCTATATGTCA 292
              |||||   |   ||||| |||||| | |||||||||| || || |  ||||||  || |||
Db        179 TTGGGAGTGCGGTCAAATGACCCTATCACAGGGGTCTCAAATGAGATATCCTGCATATCA 238

Qy        293 GGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAAT 352
                |||||||||||| ||||| ||  |||| || ||||||||  ||||||  |  ||   ||||
Db        239 AATATTTACATTATGATTCATAGTAGTACCAGAATTACAGTTATGAAGTTACA---AAAT 295

Qy        353 AATTTTATGATTGAAGGTCACCACAACATG 382
              ||||||||   |||  ||||||||||||||
Db        296 AATTTTATAGCTGAGAGTCACCACAACATG 325


RESULT 3
US-09-338-907-183
; Sequence 183, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: PROSTATE CANCER GENE
;  FILE REFERENCE: GENSET.18CP1CP
;  CURRENT APPLICATION NUMBER: US/09/338,907
;  CURRENT FILING DATE: 1999-06-23
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 09/218,207
;  EARLIER FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 183
;   LENGTH: 37950
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5259..5328
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 12675..12791
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14621..14710
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19822..19912
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 21789..21950
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 23387..23510
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25520..26016
;   OTHER INFORMATION: exon8
US-09-338-907-183

  Query Match            3.2%;  Score 74.8;  DB 3;  Length 37950;
  Best Local Similarity   80.0%;  Pred. No. 6.5e-12;
  Matches   88;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        276 ATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTA 335
              | |||||||   | |||| || |||||||| || | |||||| ||||||||  |||| ||
Db       9570 ACCTATCCTGCTTATCAGATAGTTACATTATGAATTGTAACAGCAGCAAAATCACAGTTA 9629

Qy        336 TGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              | |||| |||  |||||||||||| |||| ||||||||  ||| |||||
Db       9630 CGCAATATCAACAAAATAATTTTATGGTTGAGGGTCACCATAACGTGAGG 9679


RESULT 4
US-09-218-207-183
; Sequence 183, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Prostate cancer gene
;  FILE REFERENCE: GENSET.018CP1
;  CURRENT APPLICATION NUMBER: US/09/218,207
;  CURRENT FILING DATE: 1998-12-22
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
```

```
; SEQ ID NO 183
;   LENGTH: 37950
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5259..5328
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 12675..12791
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14621..14710
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19822..19912
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 21789..21950
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 23387..23510
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25520..26016
;   OTHER INFORMATION: exon8
US-09-218-207-183

  Query Match             3.2%;  Score 74.8;  DB 4;  Length 37950;
  Best Local Similarity   80.0%;  Pred. No. 6.5e-12;
  Matches   88;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        276 ATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTA 335
              | |||||||   | |||| || |||||||| || | ||||||| |||||||| |||| ||
Db       9570 ACCTATCCTGCTTATCAGATAGTTACATTATGAATTGTAACAGCAGCAAAATCACAGTTA 9629

Qy        336 TGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              | |||| |||  |||||||||||| |||| |||||||| ||| |||||
Db       9630 CGCAATATCAACAAAATAATTTTATGGTTGAGGGTCACCATAACGTGAGG 9679


RESULT 5
US-08-135-511-35/c
; Sequence 35, Application US/08135511
; Patent No. 5558999
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, John
;    TITLE OF INVENTION:  Cholesterol 7a-Hydroxylase Gene
;    TITLE OF INVENTION:  Regulatory Elements and Methods for Using Them
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/135,511
;      FILING DATE:  13-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SANDERCOCK, Colin G.
;      REGISTRATION NUMBER:  31,298
;      REFERENCE/DOCKET NUMBER:  18748/175
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10614 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-135-511-35

  Query Match             3.2%;  Score 74.6;  DB 1;  Length 10614;
  Best Local Similarity   64.0%;  Pred. No. 3.5e-12;
  Matches  130;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    1;

Qy        195 GATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGG----GAATCAAA 250
              . || |||| | |||   |||||||||| ||  | | |||| | | |||    || |||||
Db       7215 GAATCTAAGTCAGTGCTTCTCAACCTGTGGATCACAACCCATGTAGGGGGTAGAGTCAAA 7156

Qy        251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
              ||||||||   |||           | |||  | |  ||| ||||| | ||| ||||
Db       7155 TGACCCTTTCACAGGGGTCACCTAAGACCATCAGAAAACACAGATATTTGCTTTATGATT 7096

Qy        311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
              | | |||| |||||||||| || | |||| |||||||||| | |||  | |  ||
Db       7095 CATGACAGAAGCAAAATTATAGTTGTGAAGTAGCAATGAAAATAATTTCACAGTTGGGGG 7036

Qy        371 CACCACAACATGAGGCCGCCACA 393
              |||||||| |||  |  | |
Db       7035 TCACACAACATCAGGAAGGTAGA 7013


RESULT 6
US-08-187-453-35/c
; Sequence 35, Application US/08187453
; Patent No. 5753431
;  GENERAL INFORMATION:
;    APPLICANT:  Chiang, John
;    TITLE OF INVENTION:  Cholesterol 7a-Hydroxylase Gene
;    TITLE OF INVENTION:  Regulatory Elements and Transcription Factors
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,453
;      FILING DATE:  28-JAN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/135,488
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/135,511
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/135,510
;      FILING DATE:  13-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SANDERCOCK, Colin G.
;      REGISTRATION NUMBER:  31,298
;      REFERENCE/DOCKET NUMBER:  18748/188
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10614 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-187-453-35

  Query Match             3.2%;  Score 74.6;  DB 1;  Length 10614;
  Best Local Similarity   64.0%;  Pred. No. 3.5e-12;
  Matches  130;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    1;

Qy        195 GATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGG----GAATCAAA 250
              || |||| | |||   |||||||||| ||  | | |||| | | |||    || |||||
Db       7215 GAATCTAAGTCAGTGCTTCTCAACCTGTGGATCACAACCCATGTAGGGGGTAGAGTCAAA 7156

Qy        251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
              ||||||||    |||          | |||  | |   ||| ||||| | ||| ||||
Db       7155 TGACCCTTTCACAGGGGTCACCTAAGACCATCAGAAAACACAGATATTTGCTTTATGATT 7096

Qy        311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
              | | |||| |||||||||| || | |||| ||||||||||  | |||   | |   ||
Db       7095 CATGACAGAAGCAAAATTATAGTTGTGAAGTAGCAATGAAAATAATTTCACAGTTGGGGG 7036

Qy        371 CACCACAACATGAGGCCGCCACA 393
                 |||||||| |||  |  | |
Db       7035 TCACACAACATCAGGAAGGTAGA 7013


RESULT 7
US-08-882-164D-38
; Sequence 38, Application US/08882164D
; Patent No. 6306624
;  GENERAL INFORMATION:
;    APPLICANT:  Petkovich, P. Martin, White, Jay A.,
;    APPLICANT:  Beckett, Barbara R., Jones, Glenville
;    TITLE OF INVENTION:  Retinoid Metabolizing Protein
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Blake, Cassels & Graydon
;      STREET:  Box 25, Commerce Court West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5L 1A9
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3 1/2 inch, 1.4 Mb storage
;      COMPUTER:  COMPAQ, IBM PC compatible
;      OPERATING SYSTEM:  MS-DOS 5.1
;      SOFTWARE:  WORD PERFECT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/882,164D
;      FILING DATE:  June 25, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/667,546
;      FILING DATE:  June 21, 1996
;      APPLICATION NUMBER:  08/724,466
;      FILING DATE:  October 1, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunt, John C.
;      REGISTRATION NUMBER:  36,424
;      REFERENCE/DOCKET NUMBER:  50767/00010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 863-4344
;      TELEFAX:  (416) 863-2653
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4164 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-882-164D-38

  Query Match             3.1%;  Score 73.4;  DB 4;  Length 4164;
  Best Local Similarity   70.2%;  Pred. No. 4.7e-12;
  Matches  127;  Conservative    0;  Mismatches   51;  Indels    3;  Gaps    2;

Qy        206 AGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAGGG 265
              ||   || |||||||| ||| |  ||||||||   ||      ||    || || | | ||
Db       1252 AGAGCTTTTCAACCTGTGGGTCGTGACCCCTTCACGGAGCCAAACAACCCTTTCAGAAGG 1311

Qy        266 GTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAA 325
              ||| | ||  | |    ||  || || | ||||||||| | ||  | ||||||||||||
Db       1312 GTCGCCTAAGAGC--ATCTGCATATCCGATATTTACATCAAGAAACATAACAGTAGCAAA 1369

Qy        326 ATTACAGGTATGAAATAGCAATGAA-ATAATTTTATGATTGAAGGTCACCACAACATGAG 384
              ||||| | |||||| ||||||  || ||||||||||  |||  ||||||||||||| |||
Db       1370 ATTACCGTTATGAAGTAGCAACAAAGATAATTTTATCGTTGGGGGTCACCACAACACGAG 1429

Qy        385 G 385
              |
Db       1430 G 1430


RESULT 8
US-09-272-496-7
; Sequence 7, Application US/09272496
; Patent No. 6245966
; GENERAL INFORMATION:
;  APPLICANT: DeGregori, James
;  TITLE OF INVENTION: Adenoviral mediated gene transfer into lymphocytes
;  FILE REFERENCE: 90-98
;  CURRENT APPLICATION NUMBER: US/09/272,496
;  CURRENT FILING DATE: 1999-03-19
;  EARLIER APPLICATION NUMBER: US 60/092782
;  EARLIER FILING DATE: 1998-07-14
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 4072
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-272-496-7

  Query Match             2.9%;  Score 68.6;  DB 3;  Length 4072;
  Best Local Similarity   69.5%;  Pred. No. 1.4e-10;
  Matches  141;  Conservative    0;  Mismatches   49;  Indels   13;  Gaps    3;

Qy        195 GATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGAC 254
              ||   ||| |||   |||||||| ||| ||| | | |||||||| ||| |  |||| |
Db         50 GACCGTAGAGCACAGGTTCTCAATCTGTGGGTCGCAACCCCTTT-GGGCATCCAAAAGCA 108

Qy        255 CCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACA----------T 303
              | |||| |  |||  |||||||  |||||  ||| |||| |||||| |          |
Db        109 CTTTTAGACAGGTTGCATATCAGATATCCAGTATATCAGATATTTATATTTAAAATTATT 168

Qy        304 TACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATG-AAATAATTTTATGA 362
              ||  | ||||| |||| |||| ||||| | ||  || |||||| | ||||||||||||
Db        169 TAAAAGTCGTAGCAGTGGCAAGATTACGGTTACAAAGTAGCAACGAAAATAATTTTATGC 228

Qy        363 TTGAAGGTCACCACAACATGAGG 385
              |||   |||| |||| |||||||
Db        229 TTGGGAGTCATCACAGCATGAGG 251


RESULT 9
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
```

```
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  208:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-208

  Query Match             2.9%;  Score 68;  DB 3;  Length 16442;
  Best Local Similarity   66.2%;  Pred. No. 5e-10;
  Matches   98;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        238 TGGGGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTAT 297
              ||   ||   ||  |  | || |||||| | ||| ||||   ||||| ||| |||| ||
Db       6043 TGATTGATCAAATGGCTCTTTCACAGGGTTTACAAATCAAACATCCTGTATATCAGATAG 5984

Qy        298 TTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTT 357
              |||||| || |||| |||| || ||||||||||||  | ||| |||||| | ||  | |
Db       5983 TTACATCACAATTCATAACGGTGGCAAAATTACAGTCAAGAAGTAGCAAAGGAAATAATG 5924

Qy        358 TATGATTGAAGGTCACCACAACATGAGG 385
              |     |    | || |||| || ||||
Db       5923 TCATGGTTGGTGCCATCACAGCACGAGG 5896


RESULT 10
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
;  GENERAL INFORMATION:
;       APPLICANT: Fu, Ying-Hui
;                  Yu, Chang-En
;                  Oshima, Junko
;                  Mulligan, John T.
;                  Schellenberg, Gerald D.
;       TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                           WERNER'S SYNDROME
;       NUMBER OF SEQUENCES: 209
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/618,166
;            FILING DATE: 17-Jul-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 240052.419C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 208:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 16442 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

  Query Match             2.9%;  Score 68;  DB 4;  Length 16442;
  Best Local Similarity   66.2%;  Pred. No. 5e-10;
  Matches   98;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        238 TGGGGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTAT 297
              ||   ||   ||  |  | || |||||| | ||| ||||   ||||| ||| |||| ||
Db       6043 TGATTGATCAAATGGCTCTTTCACAGGGTTTACAAATCAAACATCCTGTATATCAGATAG 5984

Qy        298 TTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTT 357
              |||||| || |||| |||| || ||||||||||||  | ||| |||||| | ||  | |
Db       5983 TTACATCACAATTCATAACGGTGGCAAAATTACAGTCAAGAAGTAGCAAAGGAAATAATG 5924

Qy        358 TATGATTGAAGGTCACCACAACATGAGG 385
              |     |    | || |||| || ||||
Db       5923 TCATGGTTGGTGCCATCACAGCACGAGG 5896


RESULT 11
US-08-781-891-207
; Sequence 207, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  207:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29604 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-207

  Query Match             2.8%;  Score 65.6;  DB 3;  Length 29604;
  Best Local Similarity   76.9%;  Pred. No. 4e-09;
  Matches   80;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        262 AGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAG 321
              || |  |   |  ||  ||||||  || ||||||||||||||   ||||| || |||||
Db      18748 AGCGACCCTTTCACAGATATCCTGAATATCAGGTATTTACATCGTGATTCATAGCAGTAA 18807

Qy        322 CAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTG 365
              |||||||||| |||||| || ||||||||| |||||||| |||
Db      18808 CAAAATTACAGTTATGAAGTACCAATGAAATCATTTTATGGTTG 18851


RESULT 12
US-09-618-166-207
; Sequence 207, Application US/09618166
; Patent No. 6583112
;  GENERAL INFORMATION:
;       APPLICANT: Fu, Ying-Hui
;                  Yu, Chang-En
;                  Oshima, Junko
;                  Mulligan, John T.
;                  Schellenberg, Gerald D.
;       TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                           WERNER'S SYNDROME
;       NUMBER OF SEQUENCES: 209
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/618,166
;            FILING DATE: 17-Jul-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 240052.419C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 207:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 29604 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 207:
US-09-618-166-207

  Query Match             2.8%;  Score 65.6;  DB 4;  Length 29604;
  Best Local Similarity   76.9%;  Pred. No. 4e-09;
  Matches   80;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        262 AGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAG 321
              || |  |   |  ||  ||||||  || ||||||||||||||   ||||| || |||||
Db      18748 AGCGACCCTTTCACAGATATCCTGAATATCAGGTATTTACATCGTGATTCATAGCAGTAA 18807

Qy        322 CAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTG 365
              |||||||||| |||||| || ||||||||| |||||||| |||
Db      18808 CAAAATTACAGTTATGAAGTACCAATGAAATCATTTTATGGTTG 18851


RESULT 13
US-08-380-403A-4/c
; Sequence 4, Application US/08380403A
; Patent No. 5831024
;  GENERAL INFORMATION:
;    APPLICANT:  MINATO, Nagahiro
;    APPLICANT:  HATTORI, Masakazu
;    APPLICANT:  HIROSHI, Kubota
;    APPLICANT:  MASATSUGU, Maeda
;    TITLE OF INVENTION:  SPA-1 PROTEIN AND GENE CODING THEREFOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/380,403A
;      FILING DATE:  30-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/325,909
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-279712
;      FILING DATE:  20-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-139513
;      FILING DATE:  30-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WEGNER, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  53466/128/AAOK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6645 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(904..1015, 1356..1459, 1726..1883, 2009
;      LOCATION:  ..2618, 2890..3164, 4291..4509, 4598..4709, 4795
;      LOCATION:  ..4903, 5017..5117, 5200..5255, 5447..5525, 5598
;      LOCATION:  ..5741)
US-08-380-403A-4

  Query Match             2.8%;  Score 65;  DB 2;  Length 6645;
```

```
  Best Local Similarity   72.3%;  Pred. No. 2.5e-09;
  Matches  112;  Conservative    0;  Mismatches   40;  Indels    3;  Gaps    2;

Qy        233 CCCTTTGGGGGAATCAAACGACCCTTTACAGGGGTCACAT--ATCATCTATCCTATATGT 290
              |||||  | ||    ||  | || || |||||||||| |||  | ||   | ||| | |||
Db        554 CCCTTCTGAGGGTCTAACGGCCCATTCACAGGGGTCGCATGCAGCAGAGACCCTGTGTGT 495

Qy        291 CAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
              ||||| ||||   ||| |||| |||  |||||| |||||||| |||||| | ||||||||
Db        494 CAGGTGTTTATGGTACAATTCCTAAGGGTAGCAGAATTACAGTTATGAAGTGGCAATGAA 435

Qy        351 ATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              |  | ||| ||| | | |||||||||| |||| ||
Db        434 AATAATTTTTGA-TCAGGGTCACCACACCATGTGG 401


RESULT 14
US-08-895-628-4/c
; Sequence 4, Application US/08895628
; Patent No. 5998585
;  GENERAL INFORMATION:
;    APPLICANT:  MINATO, Nagahiro
;    APPLICANT:  HATTORI, Masakazu
;    APPLICANT:  HIROSHI, Kubota
;    APPLICANT:  MASATSUGU, Maeda
;    TITLE OF INVENTION:  SPA-1 PROTEIN AND GENE CODING THEREFOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/895,628
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/380,403
;      FILING DATE:  30-JAN-1995
;      APPLICATION NUMBER:  US 08/325,909
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-279712
;      FILING DATE:  20-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-139513
;      FILING DATE:  30-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WEGNER, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  53466/128/AAOK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6645 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(904..1015, 1356..1459, 1726..1883, 2009
;      LOCATION:  ..2618, 2890..3164, 4291..4509, 4598..4709, 4795
;      LOCATION:  ..4903, 5017..5117, 5200..5255, 5447..5525, 5598
;      LOCATION:  ..5741)
US-08-895-628-4

  Query Match             2.8%;  Score 65;  DB 2;  Length 6645;
  Best Local Similarity   72.3%;  Pred. No. 2.5e-09;
  Matches  112;  Conservative    0;  Mismatches   40;  Indels    3;  Gaps    2;

Qy        233 CCCTTTGGGGGAATCAAACGACCCTTTACAGGGGTCACAT--ATCATCTATCCTATATGT 290
              |||||  | ||    ||  | || || |||||||||| |||  | ||   | ||| | |||
Db        554 CCCTTCTGAGGGTCTAACGGCCCATTCACAGGGGTCGCATGCAGCAGAGACCCTGTGTGT 495

Qy        291 CAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
              ||||| ||||   ||| |||| |||  |||||| |||||||| |||||| | ||||||||
Db        494 CAGGTGTTTATGGTACAATTCCTAAGGGTAGCAGAATTACAGTTATGAAGTGGCAATGAA 435

Qy        351 ATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              |  | ||| ||| | | |||||||||| |||| ||
Db        434 AATAATTTTTGA-TCAGGGTCACCACACCATGTGG 401


RESULT 15
US-08-895-810D-4/c
; Sequence 4, Application US/08895810D
; Patent No. 6406886
;  GENERAL INFORMATION:
;       APPLICANT: MINATO, Nagahiro
;                  HATTORI, Masakazu
;                  HIROSHI, Kubota
;                  MASATSUGU, Maeda
;       TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR
;       NUMBER OF SEQUENCES: 7
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Foley & Lardner
;            STREET: 3000 K Street, N.W., Suite 500
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: USA
;            ZIP: 20007-5109
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/895,810D
;            FILING DATE: 17-Jul-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/380,403
;            FILING DATE: 30-JAN-1995
;            APPLICATION NUMBER: US 08/325,909
;            FILING DATE: 19-OCT-1994
;            APPLICATION NUMBER: JP 6-279712
;            FILING DATE: 20-OCT-1994
;            APPLICATION NUMBER: JP 6-139513
;            FILING DATE: 30-MAY-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: WEGNER, Harold C.
;            REGISTRATION NUMBER: 25,258
;            REFERENCE/DOCKET NUMBER: 53466/128/AAOK
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (202)672-5300
;            TELEFAX: (202)672-5399
;            TELEX: 904136
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 6645 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
```

```
;             NAME/KEY:  CDS
;             LOCATION:  join(904..1015, 1356..1459, 1726..1883, 2009
;             ..2618, 2890..3164, 4291..4509, 4598..4709, 4795..4903,
;             5017..5117, 5200..5255, 5447..5525, 5598..5741)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-895-810D-4

  Query Match             2.8%;  Score 65;  DB 4;  Length 6645;
  Best Local Similarity   72.3%;  Pred. No. 2.5e-09;
  Matches  112;  Conservative    0;  Mismatches   40;  Indels    3;  Gaps    2;

Qy        233 CCCTTTGGGGGAATCAAACGACCCTTTACAGGGGTCACAT--ATCATCTATCCTATATGT 290
              |||||  | ||    ||  | || || |||||||||| |||  | ||   | ||| | |||
Db        554 CCCTTCTGAGGGTCTAACGGCCCATTCACAGGGGTCGCATGCAGCAGAGACCCTGTGTGT 495

Qy        291 CAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
              ||||| ||||   ||| |||| |||  |||||| |||||||| |||||| | ||||||||
Db        494 CAGGTGTTTATGGTACAATTCCTAAGGGTAGCAGAATTACAGTTATGAAGTGGCAATGAA 435

Qy        351 ATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              |  | ||| ||| | | |||||||||| |||| ||
Db        434 AATAATTTTTGA-TCAGGGTCACCACACCATGTGG 401


Search completed: November 18, 2003, 07:06:16
Job time : 147.266 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:33:10 ; Search time 1128.46 Seconds
(without alignments)
6829.176 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggatcctttcatgtttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2358	100.0	2358	14	US-10-005-337A-1	Sequence 1, Appli
	2	367.8	15.6	2074	14	US-10-005-337A-2	Sequence 2, Appli
	3	100.4	4.3	90650	14	US-10-175-523-80	Sequence 80, Appl
	4	88.8	3.8	659158	10	US-09-771-208-20	Sequence 20, Appl
	5	84.2	3.6	106664	14	US-10-175-523-97	Sequence 97, Appl
c	6	82	3.5	659158	10	US-09-771-208-20	Sequence 20, Appl
	7	80.6	3.4	889	12	US-10-002-631C-156	Sequence 156, App
	8	80.4	3.4	10917	12	US-10-195-963-1	Sequence 1, Appli
	9	78.2	3.3	185548	14	US-10-175-523-62	Sequence 62, Appl
	10	78	3.3	892	14	US-10-228-794-75	Sequence 75, Appl
c	11	76.8	3.3	249487	13	US-10-026-188-3	Sequence 3, Appli
	12	75.8	3.2	335	10	US-09-728-445-91	Sequence 91, Appl
c	13	75.6	3.2	5990	10	US-09-917-800A-477	Sequence 477, App
	14	74.8	3.2	37950	10	US-09-901-484A-183	Sequence 183, App
	15	74.8	3.2	37950	10	US-09-853-526-183	Sequence 183, App
c	16	71.6	3.0	171936	12	US-10-025-966A-24	Sequence 24, Appl
c	17	71.6	3.0	171936	12	US-10-265-071-24	Sequence 24, Appl
	18	69.4	2.9	1889	10	US-09-974-298-182	Sequence 182, App
	19	69.4	2.9	1889	12	US-10-240-965-158	Sequence 158, App
	20	69.4	2.9	1889	12	US-10-252-157-421	Sequence 421, App
c	21	68.8	2.9	106664	14	US-10-175-523-97	Sequence 97, Appl
c	22	68	2.9	185548	14	US-10-175-523-62	Sequence 62, Appl
	23	67.8	2.9	1325	14	US-10-084-817-217	Sequence 217, App
c	24	64	2.7	85548	14	US-10-175-523-75	Sequence 75, Appl
	25	63.8	2.7	37051	12	US-10-004-113-55	Sequence 55, Appl
	26	63.2	2.7	184	9	US-09-923-876-5218	Sequence 5218, Ap
	27	63	2.7	3240	12	US-10-125-994A-25	Sequence 25, Appl
	28	63	2.7	3537	12	US-10-125-994A-26	Sequence 26, Appl
c	29	62	2.6	7208	11	US-09-824-322B-107	Sequence 107, App
c	30	60.6	2.6	67191	12	US-10-105-612-1	Sequence 1, Appli
	31	60.4	2.6	696	9	US-09-728-446-115	Sequence 115, App
c	32	60	2.5	32069	12	US-10-004-113-7	Sequence 7, Appli
c	33	59.8	2.5	347	9	US-09-728-446-339	Sequence 339, App
	34	59.6	2.5	4109	11	US-09-866-050A-567	Sequence 567, App
	35	59.6	2.5	4109	14	US-10-152-661-567	Sequence 567, App
c	36	59.4	2.5	520	10	US-09-867-701-1380	Sequence 1380, Ap
c	37	57.6	2.4	6000	9	US-09-738-968-44	Sequence 44, Appl
c	38	57.4	2.4	893	13	US-10-044-592-3	Sequence 3, Appli
c	39	57.4	2.4	955	13	US-10-044-592-93	Sequence 93, Appl
c	40	57.4	2.4	1574	13	US-10-044-592-95	Sequence 95, Appl
c	41	56.6	2.4	1235	10	US-09-809-545A-15	Sequence 15, Appl
c	42	56.6	2.4	2120	9	US-09-801-574-29	Sequence 29, Appl
	43	56.2	2.4	74868	14	US-10-175-523-67	Sequence 67, Appl
	44	55.6	2.4	1470	14	US-10-208-304-6	Sequence 6, Appli
c	45	55.6	2.4	173808	13	US-10-003-806-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-005-337A-1
; Sequence 1, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; TITLE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2358
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-005-337A-1

Query Match 100.0%; Score 2358; DB 14; Length 2358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
```

```
Qy       181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
Db       181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240

Qy       241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
Db       241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300

Qy       301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
Db       301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360

Qy       361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
Db       361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420

Qy       421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
Db       421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480

Qy       481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540
Db       481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC 540

Qy       541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
Db       541 CAGGTCTGCTTCTTATGGGTGGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600

Qy       601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660
Db       601 CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCTAATGTTTCTGTGTCTAGGTGTCATCT 660

Qy       661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720
Db       661 CTGTGAATCGAGATCCTTGGCCTTGCTTGAATTAGGGAGGCACAAAATACTCAGAGATTC 720

Qy       721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780
Db       721 AAGACTGCTCAGCAGCCCAGAGTCCTTCCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCC 780

Qy       781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840
Db       781 TTAGCTCTGAGTCAGGCCTGGAACAAACGGCCACAGGAATGAGAAAAGCTGCCATAGCTG 840

Qy       841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900
Db       841 CTTGTCACTTCAAGAGGTCAAAGAAAATAGTGTTAACCATGAAAACGAGAAGACCAACAG 900

Qy       901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960
Db       901 TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTAGGAGAGGGGAACC 960

Qy       961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020
Db       961 CACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGCAATGTCTTGTACTGAAGATTCT 1020

Qy      1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080
Db      1021 AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGGTTCTTACCCCATGTTCA 1080

Qy      1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140
Db      1081 TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAACAAACATACACAACA 1140

Qy      1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200
Db      1141 CCCTAATTAACACTTCCCTCTTCTACTGACACCCCCTTCACTCTCCTCTTTCATAAAAAA 1200

Qy      1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260
Db      1201 TAAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCCTCGAACTATAAAAAGATC 1260

Qy      1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320
Db      1261 TAAATATTTATATTTTTCACATTTTAATATCTTAGCGATGACAAGCCAGAAACAAGTATT 1320

Qy      1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380
Db      1321 TTTTGCCTCTCTCAACAGCAAAGCTTGGGGCCTTTTTGTTTCCGTGTTAGGAATAGAACA 1380

Qy      1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440
Db      1381 CGAGAGCCCCGTGTATCTAGGCAGATGCTCTATCATTAGCCCATGAGTCTCCAGCCTCAG 1440

Qy      1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500
Db      1441 ACGCACATTTTTCTCGGGCTCTCTTAAGCTTTTCCCACAGCATTGGGAAACTTTACTGAC 1500

Qy      1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560
Db      1501 AGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTGTGCATCACTTCGG 1560

Qy      1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620
Db      1561 CCCGTTTTGGGGTAGATCCTCTGATTAGCCTTCAGATTTAGAACACGGTGAGCCTGTGGT 1620

Qy      1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680
Db      1621 GCACTAATTATGGCCAGTGACACCATAGAGTCAAAGTGCATTACTGAATGCTTTCAATTT 1680

Qy      1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740
Db      1681 CTCCTAATGCTGGTACGATGGCATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG 1740

Qy      1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800
Db      1741 AGAATTCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGA 1800

Qy      1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860
Db      1801 TTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAA 1860

Qy      1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920
Db      1861 GTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGAT 1920

Qy      1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
Db      1921 ACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980

Qy      1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040
Db      1981 ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTGCTGAGCGGTGTG 2040

Qy      2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100
Db      2041 GTCACTGCCAAAGGAATGACCCTCTCACATTTCTTCCTGATTCGCATACGCCGCGGCCAG 2100

Qy      2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160
Db      2101 CTTGTCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCT 2160

Qy      2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220
Db      2161 CTGAATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATC 2220

Qy      2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280
Db      2221 TCACCCAGCCCTAGCTATATAACGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG 2280

Qy      2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340
Db      2281 GGGTTCATCCACAAGAGAGAAAAACATAGACTCGAGGTCTAGGGAGCTTGCATGCCTGCA 2340

Qy      2341 GGTCGGAGGCCACCATGG 2358
```

```
                ||||||||||||||||||
Db       2341 GGTCGGAGGCCACCATGG 2358


RESULT 2
US-10-005-337A-2
; Sequence 2, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
;  APPLICANT: BENOIT, Patrick
;  APPLICANT:  SCHWARTZ, Bertrand
;  APPLICANT:  BRANELLEC, Didier
;  APPLICANT:  CHIEN, Kenneth R.
;  TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
;  TITLE OF INVENTION:  THEM AND USES THEREOF
;  FILE REFERENCE: 03806.0530-00000
;  CURRENT APPLICATION NUMBER: US/10/005,337A
;  CURRENT FILING DATE:  2001-12-07
;  PRIOR APPLICATION NUMBER: US 60/251,582
;  PRIOR FILING DATE: 2000-12-07
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-005-337A-2

  Query Match            15.6%;  Score 367.8;  DB 14;  Length 2074;
  Best Local Similarity  76.4%;  Pred. No. 3.4e-101;
  Matches  597;  Conservative    0;  Mismatches  162;  Indels   22;  Gaps   11;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
                |||  |  |   |  ||||   |   ||  |||||   ||   |  ||  |||   ||||    |  |  |
Db       1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 1348

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
                ||||||||||||  |||  |||   ||   |||    ||||||||||||||  ||||  |    |||||
Db       1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGACATCATAAAAGAAAAGT 1405

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
                ||||||||||||||||||||  |||   |||||  |||||  |  ||||||||||   |||||
Db       1406 GCATTACTGAATGCTTTCAATGTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 1465

Qy       1718 TTTAGCTGCAGACATCACTCCAGAGAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
                ||||||  ||||||||||||||  ||||||||||||||||||  |||||||||  |      |  |
Db       1466 TTTAGC-CCAGACATCACTCCAAAGAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524

Qy       1778 CCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCCAAAGCAACACTTC 1837
                |||  |||  |  |||     |    ||  |||   |  ||||||||||||      |||    |  |
Db       1525 CCAGATCCCTTCCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCCTGGGACAAGTTTCC 1584

Qy       1838 CCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAGTGTGATG 1897
                |    |  ||  ||  ||||  |||||||||    ||||||||||||||||   |||   .|||||||
Db       1585 C---CTAAGTGAAGTGTTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641

Qy       1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAAGGAAAA 1955
                  ||  ||    |  ||||||||||||| | ||||||||||| ||||||||||||   ||||
Db       1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCATATGAAAGCTGACAAAGAA--AAAA 1698

Qy       1956 AGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGG 2015
                ||  ||||  |||||||||||||| | |||||  ||||||||||||| | ||  ||  |  | ||
Db       1699 AGGGCAGTGATGTGGTGCAATGTCAACAGACAGCTGTCCCCTGAC-TCTTGACAAATAGG 1757

Qy       2016 ATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCACATTTCTT 2075
                ||||||  ||||||||||||||  ||||  ||||   ||||||||||| ||||||||||||||||
Db       1758 ATGACTTGCATTGCTGAGCGATGTGATCACCACCAAAGGAATGGCCCTCTCACATTTCTT 1817

Qy       2076 CCTGATTCGCATACGCCGCGG-----CCAGCTTGTCATCTCCCTCTTGGGCTTCCCAGAC 2130
                ||||||||  ||||   | ||  |       |    |   ||  |||||||||   ||||||||||
Db       1818 CCTGATTCACATATTCAGCAGGGTTAGCTTGTCCTCCCCTCCCTCTTCAGCTTCCCAGAC 1877

Qy       2131 ACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGGGTG 2190
                |||  |||||||||||||||||||||||||||||||||||||  |||||  ||  |||||||  || |||
Db       1878 ACTGAGTCTGGAATGAAAATTCACCTGCCTCTGAGTTGGCTCCTAATGGGGGCGGGAGTG 1937

Qy       2191 TGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249
                |  ||||  |  ||||||||  ||||||||||||||||||  ||||  ||||||||||  |  |  |
Db       1938 TTACTTCGGTTCCCAGGTTGGAAGATTATCTCACCCGGCCCCAGCTATATAAGCTGACCG 1997

Qy       2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACATAG 2309
                ||||||||||||  |  ||||||||  |||||||  |||  |  |  ||  ||  ||||||||||||
Db       1998 GTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGATTCCTTC-CACGACAGAAAAACATAC 2056

Qy       2310 A 2310
                |
Db       2057 A 2057


RESULT 3
US-10-175-523-80
; Sequence 80, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
;   LENGTH: 90650
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-80

  Query Match            4.3%;  Score 100.4;  DB 14;  Length 90650;
  Best Local Similarity  74.7%;  Pred. No. 5.7e-18;
  Matches  139;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
                ||||  ||||  ||||||||||  |  ||  |   ||||||  |||  ||  |  ||     ||  |
Db      56512 TCTATAGCAGTGGTTCTCAACCAGTGGCTCATAACCCCTATGGTGG-CTGAATGACCCTT 56570

Qy        258 TTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACA 317
                |  |  ||  ||||||||||    ||||  |   ||  |||   ||||||||||||||||  ||||
Db      56571 TCATAGAGGTCACATATCTGATATCTTGCATATCACATATTTACATTACGATTCATAACA 56630

Qy        318 GTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACA 377
                ||||  |||  ||||||   |||||  ||||||||||||||||||||  ||    |||  ||||||
Db      56631 GTAGGAAAGTTACAGTCATGAAGTAGCAATGAAATAATTTTATGGTTTTTGGTGACCACA 56690

Qy        378 ACATGA 383
                  | ||
Db      56691 TGAGGA 56696
```

```
RESULT 4
US-09-771-208-20
; Sequence 20, Application US/09771208
; Patent No. US20020155564A1
; GENERAL INFORMATION:
;  APPLICANT: MEDRANO, JUAN
;  APPLICANT:  BRADFORD, ERIC
;  APPLICANT:  HORVAT, SIMON
;  TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
;  FILE REFERENCE: 407T-923710US
;  CURRENT APPLICATION NUMBER: US/09/771,208
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 08/999,477
;  PRIOR FILING DATE: 1997-12-29
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 659158
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (123459)..(123478)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (602466)..(602485)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (546998)..(547017)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (494715)..(494814)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (390986)..(391005)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (346860)..(346823)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (317174)..(317193)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (280353)..(280373)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (271829)..(271848)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (183872)..(183891)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (170625)..(170645)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (132680)..(132700)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n is a, c, g, or t
US-09-771-208-20

  Query Match             3.8%;  Score 88.8;  DB 10;  Length 659158;
  Best Local Similarity   74.7%;  Pred. No. 8.5e-14;
  Matches  139;  Conservative    0;  Mismatches   42;  Indels    5;  Gaps    2;

Qy        204 GCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAG 263
              |||||||||| |||| || ||| |  || | ||||||| |    ||| | || || ||||
Db     208828 GCAGCAGTTCCCAACGTGTGGGTCCAGATCTCTTTGGGAG-CCGAAAGGGCCTTTCACAG 208886

Qy        264 GGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCA 323
              ||||||||  |||  || |||  ||||    |||||| ||||| |||| || |||  |||
Db     208887 GGGTCACAGGTCAGATACCCTGAATGTTGCATATTTATATTACAATTCATAGCAGCTGCA 208946

Qy        324 AAATTACAGGTATGAAATAGCAATGAAATAATTTTA----TGATTGAAGGTCACCACAAC 379
              ||| ||||| |||||| ||||||||||||||| |||    ||   | ||||||||||| |
Db     208947 AAACTACAGCTATGAAGTAGCAATGAAATAATCTTACAGTTGGGGGGAGGTCACCACAGC 209006

Qy        380 ATGAGG 385
              ||||||
Db     209007 ATGAGG 209012


RESULT 5
US-10-175-523-97
; Sequence 97, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
;   LENGTH: 106664
;   TYPE: DNA
;   ORGANISM: Mus musculus domesticus
US-10-175-523-97

  Query Match             3.6%;  Score 84.2;  DB 14;  Length 106664;
  Best Local Similarity   68.8%;  Pred. No. 6e-13;
  Matches  130;  Conservative    0;  Mismatches   58;  Indels    1;  Gaps    1;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
              ||||   ||    |||| || ||| |||    | || |||| || |  ||||| ||||||
Db     35071 TCTAATACAATGATTCTTAATCTGTGGGGAGTGGCCTCTTTAGGAGGTTCAAAGGACCCT 35130

Qy        258 TTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACA 317
              ||    |||   |       | |||  | |    || |||||||||| ||||| ||| |
Db     35131 TTCATAGGGGTTCCATAAGACCATCGGAAAACATAGATATTTACATTATGATTCATAATA 35190

Qy        318 GTAGCAAAATTACAGGTATGAAATAGCAAT-GAAATAATTTTATGATTGAAGGTCACCAC 376
              |||||||||||||  |||| | || ||||  |||||||||||| |||  |||||||||
Db     35191 GTAGCAAAATTACATTTATGGAGTAACAATAAAAATAATTTTATGGTTGGGGGTCACCAC 35250

Qy        377 AACATGAGG 385
              ||||| | |
Db     35251 AACATAATG 35259


RESULT 6
US-09-771-208-20/c
; Sequence 20, Application US/09771208
; Patent No. US20020155564A1
; GENERAL INFORMATION:
;  APPLICANT: MEDRANO, JUAN
;  APPLICANT:  BRADFORD, ERIC
```

```
;  APPLICANT:  HORVAT, SIMON
;  TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
;  FILE REFERENCE: 407T-923710US
;  CURRENT APPLICATION NUMBER: US/09/771,208
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 08/999,477
;  PRIOR FILING DATE: 1997-12-29
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 659158
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (123459)..(123478)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (602466)..(602485)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (546998)..(547017)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (494715)..(494814)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (390986)..(391005)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (346860)..(346823)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (317174)..(317193)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (280353)..(280373)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (271829)..(271848)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (183872)..(183891)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (170625)..(170645)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (132680)..(132700)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n is a, c, g, or t
US-09-771-208-20

  Query Match             3.5%;  Score 82;  DB 10;  Length 659158;
  Best Local Similarity   71.8%;  Pred. No. 1e-11;
  Matches  135;  Conservative    0;  Mismatches   50;  Indels    3;  Gaps    2;

Qy        198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCT 257
              |||||   ||  ||||||||||  |||  || ||| |||||||  |  ||| |  |
Db     439418 TCTAGACTAGTGGTTCTCAACCTATGGGTTTCAACCTCTTTGGGAGTTTCATATCAGATA 439359

Qy        258 TTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACA 317
              ||    | ||  |||||||   |||||  || |||| ||||| ||||| ||  | |||
Db     439358 TT--CTGCGTAGCATATCAGACATCCTGCATATCAGATATTTCCATTATGACACATAATG 439301

Qy        318 GTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACA 377
              |||||||||| |||| |||||  || |||||||| |||||||| |||  ||||||| ||
Db     439300 GTAGCAAAATCACAGTTATGAGGTAACAATGAAATGATTTTATGGTTG-GGGTCACCGCA 439242

Qy        378 ACATGAGG 385
              |||||  ||
Db     439241 ACATGTGG 439234
```

```
RESULT 7
US-10-002-631C-156
; Sequence 156, Application US/10002631C
; Publication No. US20030157486A1
; GENERAL INFORMATION:
;  APPLICANT: Graff, Jonathon M.
;  APPLICANT:  Muenster, Matthew
;  TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
;  FILE REFERENCE: A34943 090495.0243
;  CURRENT APPLICATION NUMBER: US/10/002,631C
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: 60/300,309
;  PRIOR FILING DATE: 2001-06-21
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 156
;   LENGTH: 889
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)...(203)
;   OTHER INFORMATION: n = A, C, G or T
US-10-002-631C-156

  Query Match             3.4%;  Score 80.6;  DB 12;  Length 889;
  Best Local Similarity   74.8%;  Pred. No. 2.5e-13;
  Matches  101;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        251 CGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATT 310
              | || |  |   || |||  |||||   ||||||  || ||   |||||||| ||||||
Db        441 CAACACCCTTTGGGCGTCCTATATCCGATATCCTGCATATCCAATATTTACATGACGATT 500

Qy        311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGT 370
              |  |||||  |||||||||||||||||||  ||||||  ||||||  |||| | |||  | |
Db        501 CACAACAGGCGCAAAATTACAGGTATGAAGTAGCAACAAAATAACTTTAGGGTTGGGGAT 560

Qy        371 CACCACAACATGAGG 385
              ||||||  ||||||||
Db        561 CACCACGACATGAGG 575


RESULT 8
US-10-195-963-1
; Sequence 1, Application US/10195963
; Publication No. US20030167488A1
; GENERAL INFORMATION:
;  APPLICANT: Roberds, Steven L
;  APPLICANT:  Huff, Rita M
;  TITLE OF INVENTION: MICE HETEROZYGOUS FOR WFS1 GENE AS MOUSE MODELS FOR
;  TITLE OF INVENTION:  DEPRESSION
;  FILE REFERENCE: 28341/6284.N
;  CURRENT APPLICATION NUMBER: US/10/195,963
;  CURRENT FILING DATE:  2002-07-16
;  PRIOR APPLICATION NUMBER: US/09/871,107
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 10917
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY:
;   LOCATION:
;   OTHER INFORMATION: m=a or c;  w=a or t;  y=c or t;  n=g, a, c or t
US-10-195-963-1

  Query Match             3.4%;  Score 80.4;  DB 12;  Length 10917;
  Best Local Similarity   70.8%;  Pred. No. 1.7e-12;
  Matches  136;  Conservative    0;  Mismatches   51;  Indels    5;  Gaps    2;
```

```
Qy        197 TTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCC 256
              | | ||| |||  |||||||||||| ||| | | ||||||||||||    |  || |||
Db      10340 TGCGAGGACAGAGGTTCTCAACCTGTGGGGCACAACCCCTTTGGGG--TTTGAATGACTT 10397

Qy        257 TTTACAGGGGTCACATATCA---TCTATCCTATATGTCAGGTATTTACATTACGATTCGT 313
              || ||||||||  | |  ||    | |||  | |   |||||||||||||| | |||| |
Db      10398 TTCACAGGGGTTGCTTGCCAAAGACCATCAGAAAACACAGGTATTTACATTCCAATTCAT 10457

Qy        314 AACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCAC 373
              |||||||  |||||||||| |||| | ||||||||||  | |||     |   ||| ||
Db      10458 GACAGTAGTAAAATTACAGTTATGCAGTAGCAATGAAAATAATTTTATGGTTGGGGTGAC 10517

Qy        374 CACAACATGAGG 385
              |||| ||||| |
Db      10518 CACATCATGAAG 10529


RESULT 9
US-10-175-523-62
; Sequence 62, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
;  APPLICANT: Brockman, Jeffrey
;  APPLICANT:  Evans, David
;  APPLICANT:  Hook, Derek
;  APPLICANT:  Klimczak, Leszek
;  APPLICANT:  Laeng, Pascal
;  APPLICANT:  Palfreyman, Michael
;  APPLICANT:  Rajan, Prithi
;  TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
;  FILE REFERENCE: 3235/1J795-US3
;  CURRENT APPLICATION NUMBER: US/10/175,523
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/299,151
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: US 60/317,828
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/325,150
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/333,047
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: US 60/349,936
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/361,834
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62
;   LENGTH: 185548
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-175-523-62

  Query Match             3.3%;  Score 78.2;  DB 14;  Length 185548;
  Best Local Similarity   77.2%;  Pred. No. 6.2e-11;
  Matches   95;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        263 GGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGC 322
              | ||||||| ||||  ||||||  || |||  ||||||||||  |||| |||||  ||||
Db      55602 GAGGTCACACATCAGATATCCTCCATATCAAATATTTACATTATAATTCATAACACTAGC 55661

Qy        323 AAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATG 382
              |||||||  || ||  || ||||||  |||||||||||||  |    || || ||||| |||
Db      55662 AAAATTAAAGTTACAAAGTAGCAACAAAATAATTTTATGGTAAGGGGCCATCACAATATG 55721

Qy        383 AGG 385
              |||
Db      55722 AGG 55724


RESULT 10
US-10-228-794-75
; Sequence 75, Application US/10228794
; Publication No. US20030027198A1
;  GENERAL INFORMATION:
;       APPLICANT: VANDERBILT UNIVERSITY
;                  305 Kirkland Hall
;                  Nashville, TN 37240
;       TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN VIRAL
;                           INFECTION
;       NUMBER OF SEQUENCES: 83
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Needle & Rosenberg, P.C.
;            STREET: 127 Peachtree Street, Suite 1200
;            CITY: Atlanta
;            STATE: Georgia
;            COUNTRY: USA
;            ZIP: 30303-1811
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/228,794
;            FILING DATE: 27-Aug-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/171,209
;            FILING DATE: 08-Mar-1999
;            APPLICATION NUMBER: PCT/US97/06067
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Selby, Elizabeth
;            REGISTRATION NUMBER: 38,298
;            REFERENCE/DOCKET NUMBER: 22000.0061/P
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 404 688 0770
;            TELEFAX: 404 688 9880
;  INFORMATION FOR SEQ ID NO: 75:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 892 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-10-228-794-75

  Query Match             3.3%;  Score 78;  DB 14;  Length 892;
  Best Local Similarity   70.5%;  Pred. No. 1.5e-12;
  Matches  148;  Conservative    0;  Mismatches   55;  Indels    7;  Gaps    3;

Qy        177 CTCTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCT 236
              | | | ||      || |  | ||||| |||    ||||  ||||| | |   | ||||||
Db        119 CACGTACATAGATAGTCAAAATCTAGAGCACTGTTTCTATACCTGTGAGTTGCAACCCCT 178

Qy        237 TTGGG---GGAATCAAACGACCCT-TTACAGGGGTCACATATCATCTATCCTATATGTCA 292
              |||||    |   ||||| |||||| | ||||||||| || || |  ||||||  || |||
Db        179 TTGGGAGTGCGGTCAAATGACCCTATCACAGGGGTCTCAAATGAGATATCCTGCATATCA 238

Qy        293 GGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAAT 352
              |||||||||||| ||||| ||  |||| || |||||||| |||||| |  ||   ||||
Db        239 AATATTTACATTATGATTCATAGTAGTACCAGAATTACAGTTATGAAGTTACA---AAAT 295

Qy        353 AATTTTATGATTGAAGGTCACCACAACATG 382
              ||||||||   |||  ||||||||||||||
Db        296 AATTTTATAGCTGAGAGTCACCACAACATG 325


RESULT 11
US-10-026-188-3/c
; Sequence 3, Application US/10026188
; Publication No. US20020164645A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Zuker, Charles S.
;  APPLICANT:  Zhang, Yifeng
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
;  TITLE OF INVENTION:  Ion Channel
;  FILE REFERENCE: 02307E-114910US
;  CURRENT APPLICATION NUMBER: US/10/026,188
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: US 60/259,379
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 249487
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   OTHER INFORMATION: mouse genomic region containing ltrpc5
US-10-026-188-3

  Query Match             3.3%;  Score 76.8;  DB 13;  Length 249487;
  Best Local Similarity   70.2%;  Pred. No. 2e-10;
  Matches  146;  Conservative    0;  Mismatches   57;  Indels    5;  Gaps    3;

Qy        179 CTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTT 238
              | | |||  |   |  |  ||||  ||||  ||||||||||| ||     |||||||||
Db     238599 CCTACAGCCTTTCCTAGCCTCTACAGCAGTGGTTCTCAACCTGTGGTTTGAGACCCCTTT 238540

Qy        239 GGGGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATT 298
                || | ||| | ||||||||  | |||           | |||  | |    || ||||
Db     238539 --GGAAGTCATATGACCCTTTCAAAGGGCTTGCCTAAGACCATC--AGAAAATAGATATT 238484

Qy        299 TACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAAT-GAAATAATTT 357
              ||||||| ||||| || |||||||||||||| || |||||| |||||||  |||||||||
Db     238483 TACATTATGATTCATAGCAGTAGCAAAATTATAGTTATGAAGTAGCAATAAAAATAATTT 238424

Qy        358 TATGATTGAAGGTCACCACAACATGAGG 385
              | || ||| || ||||||||||||||||
Db     238423 TGTGGTTGGAGATCACCACAACATGAGG 238396


RESULT 12
US-09-728-445-91
; Sequence 91, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
;  TITLE OF INVENTION:  Animals
;  FILE REFERENCE: LEX-0102-USA
;  CURRENT APPLICATION NUMBER: US/09/728,445
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,358
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 891
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 91
;   LENGTH: 335
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(335)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-445-91

  Query Match             3.2%;  Score 75.8;  DB 10;  Length 335;
  Best Local Similarity   81.8%;  Pred. No. 3.6e-12;
  Matches  112;  Conservative    0;  Mismatches   22;  Indels    3;  Gaps    2;

Qy        210 GTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTT-ACAGGGGTC 268
              |||||||||||| |||   | |||||||| |||  |||||||||||||| ||| | ||
Db        201 GTTCTCAACCTGTGGGTTACAACCCCTTTGAGGG--TCAAACGACCCTTTCACAAGAGTT 258

Qy        269 ACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATT 328
              || ||| | |||||||  ||||||| |||||||||||| ||   |||||| |||||||||
Db        259 ACGTATTAGCTATCCTGGATGTCAGATATTTACATTACAATGTATAACAGCAGCAAAATT 318

Qy        329 ACAGGTATGAAATAGCA 345
              |||| |||||| |||||
Db        319 ACAGTTATGAAGTAGCA 335


RESULT 13
US-09-917-800A-477/c
; Sequence 477, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 477
;   LENGTH: 5990
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AB009636
US-09-917-800A-477

  Query Match             3.2%;  Score 75.6;  DB 10;  Length 5990;
  Best Local Similarity   66.7%;  Pred. No. 3.3e-11;
  Matches  124;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy        196 ATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACC 255
              | |||| | | |  ||||||||||   | | |  ||||| || ||      |||    ||
Db       4978 ACTCTAAGCCTGTGGTTCTCAACCCATGAGTCATGACCCTTTGGGATTGTCCAATGACCC 4919

Qy        256 CTTT---ACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCG 312
              |||   |||||||||  |||  |  |||||| ||  |||| ||| |||||| | ||||
Db       4918 TTTTATCACAGGGGTCGAATAGTAGGTATCCTGTAGATCAGATATATACATTGCAATTCA 4859

Qy        313 TAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCA 372
              |||| ||  |||||||||||| | ||||||||||  |||  | |||     |  |||||
Db       4858 TAACCGTGTCAAAATTACAGTTGTGAAATAGCAACTAAAATAATTTTGTGGTTTGGGTCA 4799

Qy        373 CCACAA 378
              | ||||
```

```
Db       4798 CTACAA 4793


RESULT 14
US-09-901-484A-183
; Sequence 183, Application US/09901484A
; Patent No. US20020119460A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  TITLE OF INVENTION: Prostate Cancer Gene
;  FILE REFERENCE: GEN-T111XC3D2
;  CURRENT APPLICATION NUMBER: US/09/901,484A
;  CURRENT FILING DATE:  2001-07-09
;  PRIOR APPLICATION NUMBER: US 08/996,306
;  PRIOR FILING DATE: 1997-12-22
;  PRIOR APPLICATION NUMBER: US 60/099,658
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: US 09/218,207
;  PRIOR FILING DATE: 1998-12-22
;  PRIOR APPLICATION NUMBER: US 09/338,907
;  PRIOR FILING DATE: 1999-06-23
;  PRIOR APPLICATION NUMBER: US 09/853,526
;  PRIOR FILING DATE: 2001-05-11
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 183
;   LENGTH: 37950
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (616)..(616)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (1552)..(1552)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2809)..(2809)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2821)..(2824)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2826)..(2826)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2828)..(2829)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2831)..(2833)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (2835)..(2835)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (5259)..(5328)
;   OTHER INFORMATION: exon 2
;   NAME/KEY: misc_feature
;   LOCATION: (6247)..(6247)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8667)..(8667)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8669)..(8669)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8671)..(8679)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8681)..(8685)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8687)..(8687)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8689)..(8693)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (8698)..(8698)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (9062)..(9062)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (12298)..(12298)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (12675)..(12791)
;   OTHER INFORMATION: exon 3
;   NAME/KEY: misc_feature
;   LOCATION: (14621)..(14710)
;   OTHER INFORMATION: exon 4
;   NAME/KEY: misc_feature
;   LOCATION: (17578)..(17578)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (17639)..(17639)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (17707)..(17707)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (17867)..(17867)
;   OTHER INFORMATION: n = a, c, g, or t.
;   NAME/KEY: misc_feature
;   LOCATION: (19822)..(19912)
;   OTHER INFORMATION: exon 5
;   NAME/KEY: misc_feature
;   LOCATION: (21789)..(21950)
;   OTHER INFORMATION: exon 6
;   NAME/KEY: misc_feature
;   LOCATION: (23387)..(23510)
;   OTHER INFORMATION: exon 7
;   NAME/KEY: misc_feature
;   LOCATION: (25520)..(26016)
;   OTHER INFORMATION: exon 8
;   NAME/KEY: misc_feature
;   LOCATION: (37931)..(37931)
;   OTHER INFORMATION: n = a, c, g, or t.
US-09-901-484A-183

  Query Match             3.2%;  Score 74.8;  DB 10;  Length 37950;
  Best Local Similarity   80.0%;  Pred. No. 2.2e-10;
  Matches   88;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        276 ATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTA 335
              | ||||||| | |||| || |||||||| || | ||||||| |||||||| |||| ||
Db       9570 ACCTATCCTGCTTATCAGATAGTTACATTATGAATTGTAACAGCAGCAAAATCACAGTTA 9629

Qy        336 TGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              | |||| ||| |||||||||||| |||| |||||||| ||| |||||
Db       9630 CGCAATATCAACAAAATAATTTTATGGTTGAGGGTCACCATAACGTGAGG 9679


RESULT 15
US-09-853-526-183
; Sequence 183, Application US/09853526
; Patent No. US20020165345A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Ilya, Chumakov
```

```
;  APPLICANT:  Bougueleret, Lydie
;  TITLE OF INVENTION: PROSTATE CANCER GENE
;  FILE REFERENCE: GENSET.18CP1CP
;  CURRENT APPLICATION NUMBER: US/09/853,526
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: 09/338,907
;  PRIOR FILING DATE: 1999-06-23
;  PRIOR APPLICATION NUMBER: 08/996,306
;  PRIOR FILING DATE: 1997-12-22
;  PRIOR APPLICATION NUMBER: 60/099,658
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 09/218,207
;  PRIOR FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 183
;   LENGTH: 37950
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5259..5328
;   OTHER INFORMATION: exon2
;   NAME/KEY: exon
;   LOCATION: 12675..12791
;   OTHER INFORMATION: exon3
;   NAME/KEY: exon
;   LOCATION: 14621..14710
;   OTHER INFORMATION: exon4
;   NAME/KEY: exon
;   LOCATION: 19822..19912
;   OTHER INFORMATION: exon5
;   NAME/KEY: exon
;   LOCATION: 21789..21950
;   OTHER INFORMATION: exon6
;   NAME/KEY: exon
;   LOCATION: 23387..23510
;   OTHER INFORMATION: exon7
;   NAME/KEY: exon
;   LOCATION: 25520..26016
;   OTHER INFORMATION: exon8
US-09-853-526-183

  Query Match             3.2%;  Score 74.8;  DB 10;  Length 37950;
  Best Local Similarity   80.0%;  Pred. No. 2.2e-10;
  Matches   88;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        276 ATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTA 335
              | |||||||   | |||| || |||||||| || | ||||||| |||||||| |||| ||
Db       9570 ACCTATCCTGCTTATCAGATAGTTACATTATGAATTGTAACAGCAGCAAAATCACAGTTA 9629

Qy        336 TGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGG 385
              | |||| |||  |||||||||||||| |||| |||||||| ||| |||||
Db       9630 CGCAATATCAACAAAATAATTTTATGGTTGAGGGTCACCATAACGTGAGG 9679


Search completed: November 18, 2003, 07:42:01
Job time : 1143.46 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:10 ; Search time 4918.71 Seconds
(without alignments)
11651.427 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggatcctttcatgtttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	585	24.8	600	28	BH043261	BH043261 RPCI-24-3
2	505.4	21.4	586	28	AZ295534	AZ295534 RPCI-23-1
3	306.8	13.0	371	13	BY091041	BY091041 BY091041
4	269.6	11.4	353	13	BY079997	BY079997 BY079997
c 5	191.6	8.1	478	12	BI293043	BI293043 UI-R-DK0-
6	183	7.8	434	10	BF543101	BF543101 UI-R-AG1-
c 7	171	7.3	488	9	AW252440	AW252440 UI-R-BJ0-
c 8	171	7.3	511	9	AW251306	AW251306 UI-R-BJ0-
c 9	167.8	7.1	370	9	AI710529	AI710529 UI-R-AG1-
10	131.8	5.6	735	9	AU139209	AU139209 AU139209
c 11	119.6	5.1	509	10	BE627514	BE627514 uu52e08.y
c 12	116.2	4.9	538	28	AZ804735	AZ804735 2M0065I13
c 13	116	4.9	522	28	AQ480395	AQ480395 RPCI-11-2
14	114.6	4.9	518	28	AZ114168	AZ114168 RPCI-23-4
c 15	113	4.8	596	28	AZ290439	AZ290439 RPCI-23-5
16	112.4	4.8	623	10	BB470798	BB470798 BB470798
c 17	108	4.6	577	28	AZ734079	AZ734079 RPCI-24-1
18	107.8	4.6	556	28	AZ373614	AZ373614 1M0125N21
19	107.8	4.6	589	28	BH360537	BH360537 CH230-164
c 20	107.2	4.5	839	28	BZ110003	BZ110003 CH230-164
21	107	4.5	369	28	AZ600690	AZ600690 1M0418D12
c 22	106.2	4.5	4565	11	AK052932	AK052932 Mus muscu
c 23	106	4.5	3534	11	AK038950	AK038950 Mus muscu
24	104	4.4	310	28	AZ939809	AZ939809 2M0198P19
c 25	103.6	4.4	585	14	CA882248	CA882248 K0996B07-
26	103.6	4.4	834	28	BZ185566	BZ185566 CH230-340
c 27	103.2	4.4	519	28	AZ488841	AZ488841 1M0319O13
28	103.2	4.4	640	29	BZ239325	BZ239325 CH230-282
c 29	103	4.4	574	28	AZ563281	AZ563281 RPCI-23-2
30	102.6	4.4	422	10	BF412836	BF412836 UI-R-BT1-
31	102.6	4.4	810	29	BZ264838	BZ264838 CH230-315
c 32	102	4.3	364	28	AZ348698	AZ348698 1M0085D24
33	101.8	4.3	774	28	BZ189630	BZ189630 CH230-525
c 34	101.6	4.3	676	28	AZ496223	AZ496223 1M0332H21
35	101.2	4.3	594	28	AZ389177	AZ389177 1M0149M20
36	101	4.3	477	28	AZ778894	AZ778894 2M0014H04
c 37	101	4.3	596	28	AZ856151	AZ856151 2M0160A02
38	100.8	4.3	407	13	BY549634	BY549634 BY549634
39	100.6	4.3	600	12	BG803092	BG803092 0194-80 M
c 40	100.4	4.3	591	28	AZ445982	AZ445982 1M0242B09
41	100.4	4.3	787	28	BZ128535	BZ128535 CH230-298
42	100.4	4.3	803	29	BZ234545	BZ234545 CH230-427
43	100.4	4.3	2591	11	AK036783	AK036783 Mus muscu
44	100.2	4.2	1085	29	BZ223957	BZ223957 CH230-446
45	100	4.2	473	28	AZ060257	AZ060257 RPCI-23-4

ALIGNMENTS

RESULT 1
BH043261
LOCUS BH043261 600 bp DNA linear GSS 17-JUL-2001
DEFINITION RPCI-24-330A6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-330A6, genomic survey sequence.
ACCESSION BH043261
VERSION BH043261.1 GI:14825212
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-24-330A6.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC

```
          library availability, please contact Pieter de Jong
          (pdejong@mail.cho.org). Clones may be purchased from BACPAC
          Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
          page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
          Plate: 330  row: A  column: 6
          Seq primer: SP6
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-330A6"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      158 a    139 c    142 g    161 t
ORIGIN

  Query Match             24.8%;  Score 585;  DB 28;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-148;
  Matches  585;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        211 TTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAGGGGTCAC 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 TTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAGGGGTCAC 75

Qy        271 ATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTAC 330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 ATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTAC 135

Qy        331 AGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGGCCGCC 390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 AGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAGGCCGCC 195

Qy        391 ACACTGTTCTAGAGAAAAATCACCTGGGTGGGGAAAGGTTTGGGAAAGCCTTTCTGTCCA 450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 ACACTGTTCTAGAGAAAAATCACCTGGGTGGGGAAAGGTTTGGGAAAGCCTTTCTGTCCA 255

Qy        451 TTCTTCATTCTTCAAAGTGATGTGTTCACAGAAAGCCTTTCAGCTGTTCTGCTGGGGCTC 510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 TTCTTCATTCTTCAAAGTGATGTGTTCACAGAAAGCCTTTCAGCTGTTCTGCTGGGGCTC 315

Qy        511 TTAGTAAGTCTGAGTAGGAACTGTATGTACCAGGTCTGCTTCTTATGGGTGGAGCCAAGA 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 TTAGTAAGTCTGAGTAGGAACTGTATGTACCAGGTCTGCTTCTTATGGGTGGAGCCAAGA 375

Qy        571 CGCATCGTGGGTGGAGCGAAGACGCAACCTCACCTTCTAGCTCTGCATCCATAGCAAGTA 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 CGCATCGTGGGTGGAGCGAAGACGCAACCTCACCTTCTAGCTCTGCATCCATAGCAAGTA 435

Qy        631 GCCTAATGTTTCTGTGTCTAGGTGTCATCTCTGTGAATCGAGATCCTTGGCCTTGCTTGA 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCCTAATGTTTCTGTGTCTAGGTGTCATCTCTGTGAATCGAGATCCTTGGCCTTGCTTGA 495

Qy        691 ATTAGGGAGGCACAAAATACTCAGAGATTCAAGACTGCTCAGCAGCCCAGAGTCCTTCCT 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 ATTAGGGAGGCACAAAATACTCAGAGATTCAAGACTGCTCAGCAGCCCAGAGTCCTTCCT 555

Qy        751 CAAAGGAAAGGTCTCAACTCTCAGCCCCCCTTAGCTCTGAGTCAG 795
              |||||||||||||||||||||||||||||||||||||||||||||
Db        556 CAAAGGAAAGGTCTCAACTCTCAGCCCCCCTTAGCTCTGAGTCAG 600


RESULT 2
AZ295534
LOCUS       AZ295534                 586 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-105P3.TJ RPCI-23 Mus musculus genomic clone RPCI-23-105P3,
            genomic survey sequence.
ACCESSION   AZ295534
VERSION     AZ295534.1  GI:9537319
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-105P3.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 105  row: P  column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-105P3"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      141 a    148 c    154 g    143 t
ORIGIN

  Query Match             21.4%;  Score 505.4;  DB 28;  Length 586;
  Best Local Similarity   98.5%;  Pred. No. 8.6e-127;
  Matches  531;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy       1765 GTGGCACCCAGACCCATCTCCTTCCCCTCGGGCTGATTATCCCCAGAAATAGGATGTCCC 1824
              |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Db          1 GTGGCACCCAGACCCATCTTCTTCCCCTCGGGCTGATTATCCCCAG-AATAGGATGTCCC 59

Qy       1825 AAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGC 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 AAAGCAACACTTCCCAGCCAACTGGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGC 119

Qy       1885 TGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACA 1944
              |||||||||||||||||||||||||||||||||||||||||||||| |   | ||||||
Db        120 TGTAAGTGTGATGCACAGTGCTTGCATTTTCTTGATACGTTAGTCAT-TATTATCTGACA 178

Qy       1945 AAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCC 2004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 AAGAAGGAAAAAGAGCAGCGATGTGGTGCAATATTAACAGGCAGCTGTCCCCTGGCTTCC 238

Qy       2005 CGATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTC 2064
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 CGATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTC 298

Qy       2065 TCACATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTC 2124
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TCACATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTC 358

Qy       2125 CCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCA 2184
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 CCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCA 418

Qy       2185 GGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACG 2244
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACG 478

Qy       2245 GGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAA 2303
                  ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        479 GGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACATGAGAGAAAA 537


RESULT 3
BY091041
LOCUS       BY091041                371 bp    mRNA    linear   EST 07-DEC-2002
DEFINITION  BY091041 RIKEN full-length enriched, 10 days neonate heart Mus
            musculus cDNA clone K630084E21 5', mRNA sequence.
ACCESSION   BY091041
VERSION     BY091041.1  GI:26204320
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630084E21"
                     /tissue_type="heart"
                     /dev_stage="10 days neonate"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     heart"
BASE COUNT       91 a     94 c    105 g     81 t
ORIGIN

  Query Match           13.0%;  Score 306.8;  DB 13;  Length 371;
  Best Local Similarity   99.4%;  Pred. No. 1.8e-72;
  Matches  308;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2008 TACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCA 2067
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCTCA 60

Qy       2068 CATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCA 2127
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCCCA 120

Qy       2128 GACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGG 2187
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAGGG 180

Qy       2188 GTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGGC 2247
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGGC 240

Qy       2248 TGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACAT 2307
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAACAT 300

Qy       2308 AGACTCGAGG 2317
                  ||||||  ||
Db        301 AGACTCACGG 310


RESULT 4
BY079997
LOCUS       BY079997                353 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY079997 RIKEN full-length enriched, 10 days neonate heart Mus
            musculus cDNA clone K630022E19 5', mRNA sequence.
ACCESSION   BY079997
```

```
VERSION     BY079997.1  GI:26185447
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630022E19"
                     /tissue_type="heart"
                     /dev_stage="10 days neonate"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     heart"
BASE COUNT       82 a     88 c    102 g     79 t      2 others
ORIGIN

  Query Match            11.4%;  Score 269.6;  DB 13;  Length 353;
  Best Local Similarity  95.5%;  Pred. No. 2.8e-62;
  Matches  298;  Conservative    0;  Mismatches   11;  Indels    3;  Gaps    2;

Qy       2006 GATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCT 2065
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GATACGTGGGATGACTCGCATTGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCCTCT 61

Qy       2066 CACATTTCTTCCTGATTCGCATACGCCGCGGCCAGCTTGTCATCTCCCTCTTGGGCTTCC 2125
             |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db         62 CACATTTCTTCCTGATTCGNATACGCCGCGGNCAGCTTGTCATCTCCCTCTTGGGCTTCC 121

Qy       2126 CAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAG 2185
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGAATTGGCCACTGGTGGGGGCAG 181

Qy       2186 GGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGG 2245
             |||||||||||||||||||||||||||||||          |||||||| ||||||||||
Db        182 GGGTGTGACTTGGCTTCCCAGGCTGGAAGAT--ATCTCACCAGCCCTACCTATATAACGG 239

Qy       2246 GCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTCATCCACAAGAGAGAAAAAC 2305
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        240 GCTGGTGTGGAGGGGCTCCACAGGGCCAGTT-CAGGGGTTCATCCACAAGAGAGAAAAAC 298

Qy       2306 ATAGACTCGAGG 2317
             |||||||| ||
Db        299 ATAGACTCACGG 310


RESULT 5
BI293043/c
LOCUS       BI293043                478 bp    mRNA    linear   EST 19-JUL-2001
DEFINITION  UI-R-DK0-cdj-f-07-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
            UI-R-DK0-cdj-f-07-0-UI 3', mRNA sequence.
ACCESSION   BI293043
VERSION     BI293043.1  GI:14954159
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
```

```
            normalized rat heart pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 170-292,
            >URR1A#DNA/MER1_type
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-DK0-cdj-f-07-0-UI"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="UI-R-DK0"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                      library is a subtracted library derived from a mixture of
                      five individually tagged normalized rat libraries:
                      brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                      aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                      library was constructed from a mixture of equal amounts of
                      RNA from seven different developmental time-points:
                      embryonic day 17, embryonic day 19, embryonic day 21,
                      adult day 1, adult day 12, adult day 75, and adult day
                      200. (Exception: the aorta pool does not contain embryonic
                      day 17 RNA and the placenta pool contains only the three
                      embryonic stages). Each library was normalized
                      individually according to the procedure described by
                      Bonaldo, Lennon & Soares (Genome Research Genome 6:
                      791-806, 1996). For construction of the DK0 subtracted
                      library, plasmid DNA from each of the five individually
                      tagged normalized libraries was mixed in the proportions
                      specified above and electroporated into competent bacteria
                      for production of single-stranded circular DNA
                      representing the pool of libraries. Single-stranded
                      circular DNA representing these five normalized libraries
                      was then used as a tracer in a subtractive hybridization
                      with a driver (PCR amplified inserts from a plasmid DNA
                      template preparation) comprising: a) a set of about 1,000
                      arrayed clones from each of the five non-normalized
                      libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                      aorta (CW0s), and placenta (CX0s). The resulting pool of
                      approximately 5,000 clones represented about 33.3% of the
                      final driver population. A set of about 2,000 arrayed
                      clones from each of the five normalized libraries of brain
                      (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                      placenta (CX0). The resulting pool of about 10,000 clones
                      represented about 66.6% of the final driver population.
                      TAG_LIB=UI-R-DK0
                      TAG_TISSUE=rat heart pool
                      TAG_SEQ=ATAAGATAAC"
BASE COUNT      118 a    103 c    110 g    146 t      1 others
ORIGIN

  Query Match             8.1%;  Score 191.6;  DB 12;  Length 478;
  Best Local Similarity   69.5%;  Pred. No. 7.4e-41;
  Matches  372;  Conservative    0;  Mismatches   80;  Indels   83;  Gaps    5;

Qy          1 GGATCCTTTCATGTTTAACAATATCAACCCTAACCCAAGGGGAACAGCCTGCCTGACAGT 60
              ||||||| | ||||||||||||     |  ||| ||||  ||| | || ||||||||||||
Db        475 GGATCCTCTAATGTTTAACAAATGACAAGCTAGCCCA--GGGGAAAGACTGCCTGACAGT 418

Qy         61 GGCTTTGCCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA 120
              |||| |||||||||||||| || ||||||||    || |||||||||| ||||||||| |||
Db        417 GGCTNTGCCACCCATGAACACATCCTAGTCCCTCCCCTTTGTGAAATTCAGCCCATTCCA 358

Qy        121 ACACTTCTGCAAGCCCCATCCTCTACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCT 180
              |||                                    ||   | ||| | |||||||||
Db        357 ACA-----------------------------------GGTTCTGTCCCGAAGCTTCTCT 333

Qy        181 TTCAGGATCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGG 240
              |||||| |||||||||||||||||||||||||||||||||||||||
Db        332 TTCAGGGTCAGCCTGATTCTAGGGCAGCAGTTCTCAACCTGGA---------------- 290

Qy        241 GGGAATCAAACGACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTA 300
                                       |||||||||||||||||| ||| |||  |||| |
Db        289 ------------------------GTCACATATCATCTATCCTGTATATCAAATATTCA 255

Qy        301 CATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTAT 360
              ||| | || || |||||||| || |||||||| ||||  ||||||||||| |||||||
Db        254 CATCATGACTCATAACAGTAACACAATTACAGGCATGATGTAGCAATGAAATGATTTTAT 195

Qy        361 GATTGAAGGTCACCACAACATGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
              | |||| |||||||| |||||||||| |||||||||||| | | ||   ||   | | |
Db        194 GGTTGAGGGTCACCACCACATGAGGCCACCACACTGTTCTGGGGGAA---CAGTGGAGAG 138

Qy        421 GGGAAAGGTTTGGGAAAGCCTTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACA 480
              ||||||| ||| | ||||| ||| || |||| || ||||||||| ||||| ||||||
Db        137 AGGAAAGGATTGAGGAAGCC-TTCCATCTGTTCTGCACTCTTCAAAGGGATGTATTCACA 79

Qy        481 GAAAGCCTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTA 535
              ||||||||||   | ||| || |||||||||||||| |||||| || ||||| |
Db         78 GAAAGCCTTTGGTCGGTTTTGTTGGGGCTCTTAGTAAATCTGAGCAGAAACTGCA 24


RESULT 6
BF543101
LOCUS       BF543101                434 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  UI-R-AG1-aal-b-08-0-UI.r1 UI-R-AG1 Rattus norvegicus cDNA clone
            UI-R-AG1-aal-b-08-0-UI 5', mRNA sequence.
ACCESSION   BF543101
VERSION     BF543101.1  GI:11634214
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1790595 The following
            repetitive elements were found in this cDNA sequence: 45-167,
            >URR1A#DNA/MER1_type
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .434
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-AG1-aal-b-08-0-UI"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="UI-R-AG1"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AG1
                      library is a normalized library constructed from  13 dpc
                      rat ventricle.  The tag is a string of 6 nucleotides
```

present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa."
BASE COUNT 119 a 94 c 103 g 117 t 1 others
ORIGIN

Query Match 7.8%; Score 183; DB 10; Length 434;
Best Local Similarity 73.0%; Pred. No. 1.6e-38;
Matches 276; Conservative 0; Mismatches 96; Indels 6; Gaps 3;

```
Qy        261 CAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTA 320
              | || |||||||||||||||||| ||| |||  |||| |||| | || || |||||||||
Db         43 CTGGAGTCACATATCATCTATCCTGTATATCAAATATTCACATCATGACTCATAACAGTA 102

Qy        321 GCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACA 380
              || ||||||||| ||||  |||||||||||| |||||||| |||| ||||||||| |||
Db        103 ACACAATTACAGGCATGATGTAGCAATGAAATGATTTTATGGTTGAGGGTCACCACCACA 162

Qy        381 TGAGGCCGCCACACTGTTCTAGAGAAAAATCACCTGGGTGGGGAAAGGTTTGGGAAAGCC 440
              ||||||| |||||||||||| | | ||   ||   | | | ||||||| ||| | |||||
Db        163 TGAGGCCACCACACTGTTCTGGGGGAA---CAGTGGAGAGAGGAAAGGATTGAGGAAGCC 219

Qy        441 TTTCTGTCCATTCTTCATTCTTCAAAGTGATGTGTTCACAGAAAGCCTTTCAGCTGTTCT 500
              |||  ||  |||| || |||||||| ||||| ||||||||||||||    | ||| |
Db        220 -TTCCATCTGTTCTGCACTCTTCAAAGGGATGTATTCACAGAAAGCCTTTGGTCGGTTTT 278

Qy        501 GCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTACCAGGTCTGCTTCTTATGGGT 560
              | |||||||||||||| |||||| || ||||| |      |  | |   |||  |
Db        279 GTTGGGGCTCTTAGTAAATCTGAGCAGAAACTGCACCAGGAAATTTT--ATCTGGTTTTG 336

Qy        561 GGAGCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCTCACCTTCTAGCTCTGCATCC 620
              ||      |   || || ||||| | |||||   ||  ||   |||||| |||||||
Db        337 TGATTTTCTGCTTCTCCTGNGTGGAACCAAGACTGGACTCCATTGTCTAGCCCTGCATCT 396

Qy        621 ATAGCAAGTAGCCTAATG 638
              |||  |||  ||||||||
Db        397 GCAGCTAGTTGCCTAATG 414
```

RESULT 7
AW252440/c
LOCUS AW252440 488 bp mRNA linear EST 17-DEC-1999
DEFINITION UI-R-BJ0-adx-e-05-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone UI-R-BJ0-adx-e-05-0-UI 3', mRNA sequence.
ACCESSION AW252440
VERSION AW252440.1 GI:6596031
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 488)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized AV canal at 15 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-44, >POLY_A#Simple_repeat
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .488
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0-adx-e-05-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0 library is a subtracted library derived from the UI-R-AA1, UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_LIB=UI-R-BJ0
TAG_TISSUE=AV canal at 15 dpc
TAG_SEQ=GAAGG"
BASE COUNT 115 a 111 c 106 g 155 t 1 others
ORIGIN

Query Match 7.3%; Score 171; DB 9; Length 488;
Best Local Similarity 78.4%; Pred. No. 3.3e-35;
Matches 261; Conservative 0; Mismatches 55; Indels 17; Gaps 4;

```
Qy        879 ATGAAAACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
              ||| |||| ||||| |||||||||| ||| | ||||| ||||||||||  |||||||||
Db        349 ATGGAAACCAGAAGCCCAACAGTTGTCCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 290

Qy        939 AGAACACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGC 998
                  |||||||||||| ||||||||||||||     ||| |||||| |||||| ||||||
Db        289 ----CACTAGGAGAGGAGAACCCACGAAGGA-----TATCAGTGTGCTGGTTTCCAGGGC 239

Qy        999 AATGTCTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGT 1058
              |||||||  | | |||| |||||||||||||| ||| |||| |||| ||||||||||||||
Db        238 AATGTCTCATTCCGAAGGTTCTAGAAACACAGTTTACTGGATGAAGAGCTGAAGTGGGGT 179

Qy       1059 GGGG---GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATG 1115
              ||||   |         ||| |||||  |||||||||| |||||  |||||| ||| ||||
Db        178 GGGGTAAGAATAACAACCACGTTCACTGAAGGGTGAGGGAGGACGGACAGACATACGATG 119

Qy       1116 GCCAGCATAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
              ||||||||     ||| ||| |||||||||||||  |||||||| |||||||||  |
Db        118 GCCAGCAT-----ACAAACATAACACCCTAATTAATGCTTCCCTCTGCTACTGACACTCT 64

Qy       1176 CTTCACTCTCCTCTTTCATAAAAAATAAAAAAA 1208
              |    | ||| |||||||||||| | | ||| |
Db         63 CCCTTCACTCTTCTTTCATAAAATAAACAAACA 31
```

RESULT 8
AW251306/c
LOCUS AW251306 511 bp mRNA linear EST 17-DEC-1999
DEFINITION UI-R-BJ0-adg-f-04-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone UI-R-BJ0-adg-f-04-0-UI 3', mRNA sequence.
ACCESSION AW251306
VERSION AW251306.1 GI:6594897
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized ventricle at 13 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ0-adg-f-04-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-BJ0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
                     library is a subtracted library derived from the UI-R-AA1,
                     UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
                     UI-R-AG1 libraries. These libraries represent tissues from
                     rat atrium at 16.5 dpc, ventricle  at 16.5 dpc, AV canal
                     at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
                     canal at 15 dpc, and ventricle at 13 dpc.  The tag is a
                     string of 5-6 nucleotides present between the Not I site
                     and the oligo-dT track. The library was constructed as
                     described by Bonaldo, Lennon and Soares,  Genome Research
                     6: 791-806, 1996.
                     TAG_LIB=UI-R-BJ0
                     TAG_TISSUE=ventricle at 13 dpc
                     TAG_SEQ=CAGCGA"
BASE COUNT      127 a    120 c    113 g    151 t
ORIGIN

  Query Match           7.3%;  Score 171;  DB 9;  Length 511;
  Best Local Similarity   78.4%;  Pred. No. 3.3e-35;
  Matches  261;  Conservative    0;  Mismatches   55;  Indels   17;  Gaps    4;

Qy        879 ATGAAAACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
              ||| |||| ||||| ||||||||| ||| | ||||| ||||||||||||  |||||||||
Db        331 ATGGAAACCAGAAGGCCAACAGTTGTCCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 272

Qy        939 AGAACACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGC 998
              ||||||||||||| ||||||||||||||      ||| |||||| |||||| ||||||
Db        271 ----CACTAGGAGAGGAGAACCCACGAAGGA-----TATCAGTGTGCTGGTTTCCAGGGC 221

Qy        999 AATGTCTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGT 1058
              |||||||  | | |||| |||||||||||||| ||| |||| |||| ||||||||||||||
Db        220 AATGTCTCATTCCGAAGGTTCTAGAAACACAGTTTACTGGATGAAGAGCTGAAGTGGGGT 161

Qy       1059 GGGG---GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATG 1115
              ||||   |        ||| |||||  |||||||||| |||||  |||||| ||| ||||
Db        160 GGGGTAAGAATAACAACCACGTTCACTGAAGGGTGAGGGAGGACGGACAGACATACGATG 101

Qy       1116 GCCAGCATAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
              ||||||||     ||| ||| ||||||||||||||  |||||||||  |||||||||| |
Db        100 GCCAGCAT-----ACAAACATAACACCCTAATTAATGCTTCCCTCTGCTACTGACACTCT 46

Qy       1176 CTTCACTCTCCTCTTTCATAAAAAATAAAAAAA 1208
              |    | ||| ||||||||||||| | | ||| |
Db         45 CCCTTCACTCTTCTTTCATAAAATAAACAAACA 13


RESULT 9
AI710529/c
LOCUS       AI710529                370 bp    mRNA    linear   EST 04-JUN-1999
DEFINITION  UI-R-AG1-aal-b-08-0-UI.s1 UI-R-AG1 Rattus norvegicus cDNA clone
            UI-R-AG1-aal-b-08-0-UI 3', mRNA sequence.
ACCESSION   AI710529
VERSION     AI710529.1  GI:5000305
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized ventricle at 13 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 1-43,
            >POLY_A#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-AG1-aal-b-08-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-AG1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AG1
                     library is a normalized library constructed from  13 dpc
                     rat ventricle.  The tag is a string of 6 nucleotides
                     present  between the Not I site and the oligo-dT track.
                     The library was  constructed as described by Bonaldo,
                     Lennon and Soares,  Genome Research 6: 791-806, 1996.
                     Tissue provided by Jim Lin, Department of Biology,
                     University of Iowa.
                     TAG_LIB=UI-R-AG1
                     TAG_TISSUE=ventricle at 13 dpc
                     TAG_SEQ=CAGCGA"
```

```
BASE COUNT       75 a     82 c     80 g    132 t      1 others
ORIGIN

  Query Match             7.1%;  Score 167.8;  DB 9;  Length 370;
  Best Local Similarity   77.8%;  Pred. No. 2.2e-34;
  Matches  259;  Conservative    0;  Mismatches   57;  Indels   17;  Gaps    4;

Qy        879 ATGAAAACGAGAAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
              ||| |||| ||||| ||||||||| ||| | ||||| |||||||||||  |||||||||
Db        348 ATGGAAACCAGAAGGCCAACAGTTGTCCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 289

Qy        939 AGAACACTAGGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTTGGTTTTCAGGGC 998
                 |||||||||||| ||||||||||||||     ||| |||||| |||||   |||||
Db        288 ----CACTAGGAGAGGAGAACCCACGAAGGA-----TATCAGTGTGCTGGTTCCAAGGGC 238

Qy        999 AATGTCTTGTACTGAAGATTCTAGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGT 1058
              |||||||  | | |||| ||||||||||||| ||| |||| |||| ||||||||||||||
Db        237 AATGTCTCATTCCGAAGGTTCTAGAAACACAGTTTACTGGATGAAGAGCTGAAGTGGGGT 178

Qy       1059 GGGG---GTTCTTACCCCATGTTCATGGAAGGGTGAGTGAGGAGAGACAGATATATGATG 1115
              ||||   |        ||| |||||  |||||||||| |||||  |||||| ||| ||||
Db        177 GGGGTAAGAATAACAACCACGTTCACTGAAGGGTGAGGGAGGACGGACAGACATACGATG 118

Qy       1116 GCCAGCATAACAAACATACACAACACCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
              ||||||||     ||| ||| |||||||||||||  ||||||||| |||||||||| |
Db        117 GCCAGCAT-----ACAAACATAACACCCTAATTAATGCTTCCCTCTGCTACTGACACTCT 63

Qy       1176 CTTCACTCTCCTCTTTCATAAAAAATAAAAAAA 1208
              |    | ||| |||||||||||| | | ||| |
Db         62 CCCTTCACTCTTCTTTCATAAAATAAACAAACA 30


RESULT 10
AU139209
LOCUS       AU139209                735 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU139209 PLACE1 Homo sapiens cDNA clone PLACE1010155 5', mRNA
            sequence.
ACCESSION   AU139209
VERSION     AU139209.1  GI:11000730
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1010155"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
BASE COUNT      240 a    154 c    181 g    156 t      4 others
ORIGIN

  Query Match             5.6%;  Score 131.8;  DB 9;  Length 735;
  Best Local Similarity   83.6%;  Pred. No. 2.1e-24;
  Matches  173;  Conservative    0;  Mismatches   32;  Indels    2;  Gaps    2;

Qy       2105 TCATCTCCCTCTTGGGCTTCCCAGACACTAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
              ||  |||||||||  |||||||||||||| |||||||||||||||||||||||||||||
Db         28 TCCCCTCCCTCTTCAGCTTCCCAGACACTGAGTCTGGAATGAAAATTCACCTGCCTCTGA 87

Qy       2165 ATTGGCCACTGGTGGGGGCAGGGGTGTGACTTGGCTTCCCAGGCTGGAAGATTATCTCAC 2224
              |||||  ||  ||||||| || |||| |||| | |||||||| ||||||||||||||||
Db         88 GTTGGCTCCTAATGGGGGCGGGAGTGTTACTTCGGTTCCCAGGTTGGAAGATTATCTCAC 147

Qy       2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGG 2283
              || |||| |||||||||| | | | |||||||||||| |  ||||||||  |||||||
Db        148 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCCAGCAGGGCCAACTCCAGGGA 207

Qy       2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
              ||| | | || || |||||||||| |
Db        208 TTCCTTC-CACGACAGAAAAACATACA 233


RESULT 11
BE627514/c
LOCUS       BE627514                509 bp    mRNA    linear   EST 24-AUG-2000
DEFINITION  uu52e08.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3375590 5' similar to gb:L35933 Mouse erythrocyte protein 4.2
            (MOUSE);, mRNA sequence.
ACCESSION   BE627514
VERSION     BE627514.1  GI:9907936
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1085194
            Seq primer: -40RP from Gibco
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3375590"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_thymus_2NbMT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      150 a    110 c    125 g    123 t      1 others
ORIGIN

  Query Match             5.1%;  Score 119.6;  DB 10;  Length 509;
  Best Local Similarity   79.4%;  Pred. No. 4e-21;
```

```
  Matches  154;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy        192 CCTGATTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAAC 251
              ||  |  |||| ||||||||||||||||||| ||| ||||||||||||||||    |
Db        492 CCCTACCCTAGAGCAGCAGTTCTCAACCTGTGGGGCTCGACCCCTTTGGGGGTCGTATGA 433

Qy        252 GACCCTTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTC 311
              ||  || ||||||||    |||||  ||||||  || |||| ||||||||||| |||||
Db        432 TACT-TTCACAGGGGTAGGCTATCAGGTATCCTGCATATCAGATATTTACATTATGATTC 374

Qy        312 GTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTC 371
              |||||||||||||| ||| || | |||| ||||||||||||||  ||||| |||  ||||
Db        373 CTAACAGTAGCAAATTTATAGTTGTGAAGTAGCAATGAAATAAAGTTATGGTTGGGGGTC 314

Qy        372 ACCACAACATGAGG 385
              ||||||||||||||
Db        313 ACCACAACATGAGG 300


RESULT 12
AZ804735/c
LOCUS       AZ804735                538 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0065I13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0065I13 R, genomic survey sequence.
ACCESSION   AZ804735
VERSION     AZ804735.1  GI:12965546
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0065  row: I  column: 13
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 538.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0065I13"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      164 a    127 c    112 g    135 t
ORIGIN

  Query Match           4.9%;  Score 116.2;  DB 28;  Length 538;
  Best Local Similarity  79.9%;  Pred. No. 3.5e-20;
  Matches  151;  Conservative    0;  Mismatches   33;  Indels    5;  Gaps    1;

Qy        197 TTCTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCC 256
              ||||| ||||     |||||||||  |||  |||||||||||||||     |     ||
Db        390 TTCTAAGGCAATCTCTCTCAACCTACGGGTTTCGACCCCTTTGGG-----TATCAACCCT 336

Qy        257 TTTACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAAC 316
              || ||||||||||||||||  ||| ||  ||||||| ||||||||||||  ||| ||||
Db        335 TTCACAGGGGTCACATATCACATATTCTGCATGTCAGATATTTACATTACAGTTCATAAC 276

Qy        317 AGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCAC 376
              |||| |||||||||| | |||| ||||||||||||||||||||| ||| ||||||||||
Db        275 AGTACCAAAATTACAGTTTTGAAGTAGCAATGAAATAATTTTATGGTTGGAGGTCACCAC 216

Qy        377 AACATGAGG 385
              |||||||||
Db        215 AACATGAGG 207


RESULT 13
AQ480395/c
LOCUS       AQ480395                522 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  RPCI-11-236B22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-236B22
            , genomic survey sequence.
ACCESSION   AQ480395
VERSION     AQ480395.1  GI:4662514
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-11-236B22.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7590285"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-236B22"
```

```
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      168 a    100 c    113 g    141 t
ORIGIN

  Query Match             4.9%;  Score 116;  DB 28;  Length 522;
  Best Local Similarity   76.7%;  Pred. No. 3.9e-20;
  Matches  155;  Conservative    0;  Mismatches   45;  Indels    2;  Gaps    1;

Qy       1540 TCTCTCTGTGCATCACTTCGGCCCGTTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
              ||| | | |  | ||||  |  || |||||  ||  | || |||  ||||   | | | |
Db        202 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTATTATGCTGTT 143

Qy       1598 TTAGAACACGGTGAGCCTGTGGTGCACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
              |||||||||||| ||| |||  ||  ||||||||||||||||||| |||| |   |||||
Db        142 TTAGAACACGGTAAGCATGTCATGTGCTAATTATGGCCAGTGACATCATAAAAGAAAAGT 83

Qy       1658 GCATTACTGAATGCTTTCAATTTCTCCTAATGCTGGTACGATGGCATGTCACAGGGCCAT 1717
              |||||||||||||||||||||||||  ||||| ||||| | ||||||||||  |||||
Db          82 GCATTACTGAATGCTTTCAATTTCTTATAATGATGGTAAGGTGGCATGTCATGGGGCCTA 23

Qy       1718 TTTAGCTGCAGACATCACTCCA 1739
              ||||||  ||||||||||||||
Db          22 TTTAGCCCCAGACATCACTCCA 1


RESULT 14
AZ114168
LOCUS       AZ114168                518 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  RPCI-23-449H23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-449H23
            , genomic survey sequence.
ACCESSION   AZ114168
VERSION     AZ114168.1  GI:7774139
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-449H23.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 449  row: H  column: 23
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-449H23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      158 a     94 c    108 g    158 t
ORIGIN

  Query Match             4.9%;  Score 114.6;  DB 28;  Length 518;
  Best Local Similarity   67.2%;  Pred. No. 9.4e-20;
  Matches  162;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        145 ACAAGGTGCTCATTGGGAATTTCCTGGAGCTTCTCTTTCAGGATCAGCCTGATTCTAGGG 204
              ||| |     ||||  |||   | | | | |  |  ||     |     ||    | |  |
Db        150 ACAGGCCATGCATTTTGAAAGGCTTAGTGTTCTTGATTTTTAACACTACTTTGACGACAG 209

Qy        205 CAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTTTACAGG 264
              |||  |||||||||||| ||| | |||||||| ||||||| | |      ||  |  ||||
Db        210 CAGTGGTTCTCAACCTGTGGGTCGCGACCCCTGTGGGGGATTTACTGTCCCTCTCCCAGG 269

Qy        265 GGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAGTAGCAA 324
              |||| | |||||  ||| | | | ||||| ||| || |||||  ||  ||||| | ||||
Db        270 GGTCTCTTATCAAATATTCCACAAGTCAGATATGTATATTACAGTTTATAACAATGGCAA 329

Qy        325 AATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAACATGAG 384
              |||| ||| |||||    || ||||||||||||||  |||| ||||| ||||||||||
Db        330 AATTCCAGTTATGAGGGAGTAATGAAATAATTTTATAGTTGAGGGTCATCACAACATGAG 389

Qy        385 G 385
              |
Db        390 G 390


RESULT 15
AZ290439/c
LOCUS       AZ290439                596 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-59G22.TVB RPCI-23 Mus musculus genomic clone RPCI-23-59G22,
            genomic survey sequence.
ACCESSION   AZ290439
VERSION     AZ290439.1  GI:9532225
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-59G22.TJB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 59  row: G  column: 22
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .596
```

/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-59G22"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT 174 a 136 c 119 g 167 t
ORIGIN

Query Match 4.8%; Score 113; DB 28; Length 596;
Best Local Similarity 78.3%; Pred. No. 2.7e-19;
Matches 148; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

```
Qy        199 CTAGGGCAGCAGTTCTCAACCTGGGGGCCTCGACCCCTTTGGGGGAATCAAACGACCCTT 258
              ||| |||||  || |||||||| |||    ||||||||||||||  ||||    || ||
Db        467 CTATGGCAGTGGTCCTCAACCTGTGGGTGGTGACCCCTTTGGGGG-TTCAACAACCCTTT 409

Qy        259 TACAGGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAACAG 318
               ||||||||||||||||  | | ||   |  ||| |||||||||| |||||  |||||
Db        408 CACAGGGGTCACATATCAGATTTTCTGCGTACCAGATATTTACATTATGATTCAGAACAG 349

Qy        319 TAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCACCACAA 378
              |||||||||| ||| |||||| ||||||||||||||||| || | |  ||||| |||||
Db        348 TAGCAAAATTCCAGTTATGAAGTAGCAATGAAATAATTTTGTGGTGGGGGGTCAGCACAA 289

Qy        379 CATGAGGCC 387
              |||| || |
Db        288 CATGGGGAC 280
```

Search completed: November 18, 2003, 07:01:50
Job time : 4948.21 secs